MILITARY OPERATIONS RESEARCH SOCIETY

MORS

MORS Workshop
Test & Evaluation, Modeling and Simulation and VV&A: Quantifying the Relationship Between Testing and Simulation
15-17 October 2002
The Energy Training Complex, Kirtland AFB
Albuquerque, New Mexico

Co-Chairs:
Dr. Marion Williams, FS and Annie Patenaude

20040524 003

DISCLAIMER

This Military Operations Research Society report summarizes the proceedings of a workshop conducted over three days by experts, users and participants interested in quantifying the relationship between testing and simulation. It is not intended to be a comprehensive treatise on the subject. It reflects the major concerns, insights, thoughts and directions of the participants at the time of the workshop.

OSD Disclaimer: Review of this material does not imply Department of Defense indorsement [sic] of factual accuracy or opinion.

CAVEATS

- The Military Operations Research Society neither makes nor advocates official policy.

- Matters discussed or statements made during the workshop were the sole responsibility of the participants involved.

- The Society retains all rights regarding the final decision on the content of this workshop report.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From – To)
7 January 2004	Workshop Report	15-17 October 2002

4. TITLE AND SUBTITLE

MORS Workshop Report
Test & Evaluation, Modeling and Simulation and VV&A:
Quantifying the Relationship Between Testing and Simulation

5a. CONTRACT NUMBER

N00014-99-C-0113

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Marion Williams, FS and Annie Patenaude
Corrina Ross-Witkowski, Editor/Publisher

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Military Operations Research Society, 1703 N. Beauregard St, Suite 450, Alexandria, VA 22311

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Statement A:
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This publication is the report of the Workshop Test & Evaluation, Modeling and Simulation and VV&A: Quantifying the Relationship Between Testing and Simulation.

15. SUBJECT TERMS

16. SECURITY CLASSIFCATION: UNCLASSIFIED			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Corrina Ross-Witkowski
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	Unlimited	1-122 and Appendixes A-1-4 B-1-6	19b. TELEPHONE NUMBER (include area code) 703-933-9070 FAX: 703-933-9066

Standard Form 298 (Rev. 8-98)

CLASSIFIED BY:

DECLASSIFIED ON:

MILITARY OPERATIONS RESEARCH SOCIETY

MORS Workshop
**Test & Evaluation, Modeling and Simulation and VV&A:
Quantifying the Relationship Between Testing and Simulation**
15-17 October 2002
The Energy Training Complex, Kirtland AFB
Albuquerque, New Mexico

Co-Chairs:
Dr. Marion Williams, FS and Annie Patenaude

The Military Operations Research Society (MORS)

The purpose of the Military Operations Research Society (MORS) is to enhance the quality and effectiveness of classified and unclassified military operations research. To accomplish this purpose, the Society provides media for professional exchange and peer criticism among students, theoreticians, practitioners and users of military operations research. These media consist primarily of the traditional annual MORS Symposia (classified), their published abstracts or proceedings, special mini-symposia, workshops, colloquia and special purpose monographs and other publications. MORS publishes two quarterly periodicals, *PHALANX* and *Military Operations Research*. *PHALANX* is the MORS bulletin and *Military Operations Research* is a refereed journal. The forum provided by these media is intended to display the state of the art, to encourage consistent professional quality, to stimulate communication and interaction between practitioners and users, and to foster the interest and development of students of operations research. The Military Operations Research Society neither makes nor advocates official policy, nor does it attempt to influence the formulation of policy. Matters discussed or statements made during the course of its meetings or printed in its publications represent the positions of the individual participants and authors and not of the Society.

The Military Operations Research Society is operated by a Board of Directors consisting of 30 members, 28 of whom are elected by vote of the Board to serve a term of four years. The persons nominated for the board generally are individuals who have attained recognition and prominence in the field of military operations research, and who have demonstrated an active interest in the programs and activities of MORS. The remaining two members of the Board of Directors are the Immediate Past President who serves by right and the Executive Vice President who serves as a consequence of his position. A limited number of Advisory Directors are appointed from time to time, usually for a one-year term, to perform some particular function.

MORS is Sponsored by:

- The Deputy Under Secretary of the Army (Operations Research)
- The Director, Assessment Division, Office of the Chief of Naval Operations
- Director, Air Force Studies and Analyses Agency
- The Commanding General, Marine Corps Combat Development Command
- The Director of Force Structure, Resources and Assessment, The Joint Staff
- The Director Program Analysis and Evaluation, Office Secretary of Defense

Table of Contents

Introduction

The Military Operations Research Society (MORS) initiated a workshop on simulation validation (SIMVAL) in 1990 in response to numerous initiatives on the use of Modeling and Simulation (M&S) as a part of the acquisition process. This workshop was followed by several other workshops designed to develop a common set of terms for Verification, Validation and Accreditation (VV&A), and to define elements of VV&A that could be used to ensure that models possessed a degree of credibility – assurance that the model was adequate for addressing a specific issue. While M&S had been in use for a number of years, there was little agreement on what had to be done to ensure that the model was accurate based on internal logic checks as well as comparison with real world data.

SIMVAL produced definitions and elements for VV&A that are still in use today.

One application of M&S that has received an increased amount of attention over the past few years has been in Test and Evaluation (T&E). With increasing cost and complexity of weapon systems, and the resultant increase in the cost of testing, it has been the goal of many organizations to decrease the cost of testing through the use of M&S. While forms of simulation, such as hardware and operator in the loop simulation, have been in use in T&E for years, the test community has approached the use of constructive models with skepticism. Therefore, VV&A has been an important factor in providing confidence that a model can represent the real world to a sufficient degree to help plan tests – and achieve more efficient test programs – as well as providing an adequate supplement to testing. The use of missile fly-out models is an example of using constructive simulation to represent a part of the test environment that cannot always be included directly.

With over ten years of applications of VV&A, a new SIMVAL was proposed to assess the progress in applying VV&A to T&E applications, and to address the overall status and health of ensuring that models can adequately support T&E. The following pages summarize the 2002 SIMVAL workshop, "Test and Evaluation, Modeling and Simulation and VV&A: Quantifying the Relationship Between Testing and Simulation.

M&S Expectations for T&E and the Challenge of VV&A
Mr. Walt Hollis, FS, Deputy Undersecretary of the Army (Operations Research)

Editor's note: The following remarks were presented at the MORS Mini-Symposium on Test & Evaluation, Modeling and Simulation and VV&A at Kirtland AFB, Albuquerque, New Mexico. Other presentations from that event can be found at the MORS website at www.mors.org/completed_events.htm.

I am pleased to be here with you today as we set out to explore the issues laid forth for this conference. Dr Williams will recount for us the history of the MORS workshops on Simulation Validation. Certainly the conclusions were valid then, and are still valid now. The difficulty for the T&E community has been "how to execute these findings."

The Search for Best Practices

We know there are some best practices being used out there that can be promoted and shared universally to increase our productivity across the services and agencies. The goal of this symposium is to look at the case studies, find the best practices, and the best methodologies to enable you to execute your mission in a timely manner, effectively, and efficiently. There are two different sets of communities. There are the consumers and there are the builders. I hope we can come to a common intersection. The goal of your separate working groups will be to explore the best practices and recommend those that should be adopted universally or by groups of a particular mission focus. However, be careful not get too focused on the concept of "universality" since it may not be possible.

It is understood that we are to use models and simulations that are properly verified, validated, and accredited to support the test process. It is not only sensible, but it is also the law. What are the key issues, some of which have been around for a long time? I can think of a few before hearing the presentation. Should V&V be accomplished primarily by Subject Matter Experts (SMEs)? Who should decide the Accreditation? How rigorous need the validation be? Can models be used for both DT and OT events? What are the methods used to determine the accreditation? I submit that there are many different practices for accreditation across different agencies. What are the best practices?

It is generally agreed that we use T&E data to validate models. What are some of the difficulties associated with this? Data are scarce. It is okay at times to work with small samples, since the Lord tells us things in small doses. What must be done to enhance the process and improve the products? Indeed, resourcing is an ever-present concern. But it must not be a deterrent. It should only be a factor to consider.

> *Another issue is whether to shape a test strictly to collect data that can be used to validate a model. Our basic premise in T&E is that we use testing to increase our knowledge of the system and we use evaluation as a means to understand a system's military utility. A test and evaluation plan should be addressing these basics in addition to collecting data for validating our models. It is wasteful not to*

collect data that can be used to help validate a model, but it is equally wasteful to create test events simply to validate a model, particularly if you are involving field troops.

One of the objectives of the mini-symposium is to assess the health of VV&A, and determine if the current concept is still viable. If not, then we may need to adjust our expectations and create new objectives to improve the V&V process.

To shed some light, let me cover some history of how this test community got into the business of simulation and where it now sees itself going with simulation as we approach the analysis and testing of the transformed Army. Even with physics-based simulations, I do believe that every now and then you must get some real data from the field, from those things that are muddy and dusty. It is neither comfortable nor convenient sometimes to collect that kind of data.

The Road to Simulation by the T&E Community

Let me go back to the cold war. The Army in the late '60s had done a large experiment called "The Air Assault Division Experiment," led by then-Lieutenant General Hamilton Howze. They discovered that there was a great deal of advantage to the use of helicopters to transport squads and other Army units to battle. The speed and the agility of the helicopter via the third dimension made a difference. Somebody got a bright idea that now that we have this Cobra, maybe we can put a TOW on it to get some additional antitank capability in the Fulda Gap area, where as you know for years we expected the Russian horde to come across into Germany.

I was at CDEC at that time. (CDEC was the Combat Development Experimental Command out at Fort Ord.) Dave Hardison, my predecessor as the DUSA (OR), said to me, "I want you to run the experiment Walt, we are going to rig-up a couple of Cobras, get some of the tanks from the 41st Battalion, and we will do some experiments on how to install it, to see whether this would actually work." We also knew that we were poorly set up instrumentation wise to really do the job very well.

But we were shrewd; at least we thought we were. We said that if we do an experiment which was intriguing, but didn't really get to the full answer, the Army would produce the money for the instrumentation needed to do it right. So we proceeded. And we went to our friends at AMSAA and they provided us with the large amount of "probability of kill given a shot" data for tank-on-tank engagements and for helicopter-on-tank engagements and we went out and we did about 15 trials of this experiment under different conditions. After we got all this data, we really couldn't analyze it because after the first exchange of fire we had no way of extracting kills from the scenario. You could not analyze the data from that event onward.

But enough had been seen by the senior Army people by the little bit of data we could use. Convinced that this was probably a good idea, they invested the money in the instrumentation system which we would use for many years called the "Real-time

Casualty Assessment Technique," and it spun into the MILES system used in the National Training Centers now to make the training more realistic.

In peacetime, we train for war through the use of simulation. There is no question in my mind that some the successes we have had in our operations have been likely due to the training and the work that we had done to make the weapons that we have fielded effective, and also to understand the conditions under which they could be effective.

We have of course used simulations for many, many years on the fly-out models of the big missiles, such as the P2, the Patriot, and the Hawk, and before that the Hercules, because you could not afford to fire all the missiles that were required to get a complete mapping of their performance. So you picked out, you built a simulation, physics models basically, of the fly-out, picked some difficult shots, test those and then declare the simulation valid for other shots that are within that envelope.

The Role of Validation, Verification and Accreditation

That brings me to the point of validation, and verification, and accreditation. Validation is an important word, but I think perhaps poorly used nowadays. We started it; I am probably as much to blame as anybody. But we do want to have some basis of belief that our simulations have some creditability, that they are not pulled out of some place in the air, that they can be measured one way or another. The accreditation is a more iffy thing. It is non-technical in many ways but it is very important. It is also difficult because it is judgmental.

You have a problem, you want to use JANUS, you have the responsibility for determining whether JANUS is appropriate for the solution of that problem. That is something that a lot of people forget. As we get better and better simulations, and they look more and more real, the user must be critical as to whether the physics behind these simulations is adequate. Many users are not critical, and because they are impressed with the visualization, they believe them. Since it looks right, it is right. That's a problem.

That brings me to the point now with the Army Transformation. You know that we are buying several brigades of Strykers. They are up at Fort Lewis. They are wheeled vehicles, derivative of light Amphibious Assault Vehicle (AAV) like the Marines bought some years ago. They have eight wheels instead of six, but basically they are the same thing. We made the purchase of these vehicles based on an assessment of the ability of units equipped with this vehicle to perform a set of unit missions. There were performance specifications that had to be met such as speed made good, estimate of reliability, and certain overall level of protection. But unlike the Big 5 that the Army built in the '60s and the early '70s: the tank, the Bradley, the Patriot, the Black Hawk, and the Apache, there were no detailed military characteristics for these "eaches." Just a set of missions that the brigade had to be able to perform with the capabilities inherent in the equipment and that is how we bought the brigade.

The way we test is changed now and you have to look differently at the testing because there is no longer a key "each," which is imbedded in among other things that you already know something about and you watch to see whether that new "each" does better in its battles with its corresponding target than the last piece you had in that position. Now you can imagine that there might be situations in which the bridging equipment might be the most effective piece of equipment in accomplishing the mission that this brigade had to perform.

So the analysts had a fair problem in analyzing the alternatives and the testers had a difficult problem by not having certain parameters and criteria for each individual piece — whether the gun carriers, the missile carriers, the infantry carriers, etc. You might imagine that if the infantry went dismounted it might have made a big difference in these engagements, independent of the vehicles. So it's fun ahead! More challenges the higher you go into this problem.

To transform their approach to evaluation, the testers adopted things like JANUS and CASTFOREM, standard Army models that have been used for some years, and some other things that put together in a wrap around simulation they called, "STORM." This is a way of embedding a unit such as one of our Stryker brigades in a larger context so that orders can come in from the outside and be executed by the brigade commander, or you can use a brigade as the controller element by battalion commanders. Then you can run tests and experiments to see how these individual things, when operated as unit, perform.

Simulation for T&E with FCS and Units of Action

Simulation has become a very large part of the test community and we will see how well it does its job shortly when we begin the initial operational testing and evaluation of these new Stryker brigades. Having gone that far, we have gone on to what is called the Future Combat System (FCS), which is a collection of equipment that does not yet exist. Again, there are no requirements for "eaches," but we wish to go to a DAB on the 3rd of May in 2003. A challenge!

The only thing that we are going to have to represent as we see it now are the capabilities of these new "Units of Action" as they are called. I think, if you are familiar with what a brigade is, a Unit of Action is about the size of a brigade and has in it the assumed capability of a standard brigade. With the exception that we added the artillery and aviation lately, it looks more like a regimental combat team of World War II.

In any event, there are not going to be many physical representations of these things, to speak of, to show anybody. We are going to be able just to kick the tires and people are going to make commitments based largely on simulation and some reasonable analysis using concepts.

It is the system-of-systems environment. How do we best test it? How do we best evaluate it? How do we best analyze it? How do we best model and simulate it? What about the need to test, evaluate, analyze the soldier as a system of systems? Certainly

VV&A has a key play in ensuring we are using as correct as possible models and simulations.

We have a potential methodology called the Degraded States Taxonomy. It is the brainchild of Paul Dietz of AMSAA and others. It has been used largely for assessing battle damage by live fire. However, the concept of degraded states has applicability to a unit that has lost a part of its capability. We are going to apply this methodology to the evaluation of brigade effectiveness and look forward to using it in the analysis of FCS. A part of the success of this methodology of evaluation will depend upon the availability and development of subject matter experts who can judge mission success at brigade and battalion levels. SMEs, or Observer Controllers, the type employed at NSC and JRTC, will be required to support the evaluations of future operational tests which we now envision.

It is very different for everybody that has been involved in the Army acquisition, testing, and analysis for many years. It is a challenge every day. Those of you who are involved in the test planning for the Army Transformation are doing great work and I commend you for it.

I challenge those of you who are consumers of models and simulations to really participate in this workshop. It is structured to get everyone involved in the products so you will have plenty of opportunities. I think we have the right people here to make a difference. Optimize your time here, turn off your cell phones, your Blackberries, and your palm devices, and give your 100 percent focus to the topic at hand and contribute what you can to the discussions. What we hope to glean from the difficulties and the successes and best practices are those nuggets and recommendations for the leadership to leverage in developing policies and procedures for the future.

A Brief History of the MORS Simulation Validation Workshop Series
Dr Marion Williams, FS

In the 1990 timeframe, there was a significant increase in Modeling and Simulation (M&S) within the DoD. This increase was at least partially caused by the decrease in the Defense budget; a new acquisition process which demanded more efficiency in system design, development and testing; and, a significant increase in the capacity and capability of digital computers. With this increase in DoD M&S, there was a concern within the operations research community about the quality of the models being used, and the degree to which modelers ensured that the models were adequately addressing the questions being posed. A RAND report also questioned many aspects of M&S applications.[1]

To address the concerns regarding simulation validation, MORS initiated a series of workshops to address model Verification, Validation and Accreditation (VV&A). The purpose of the SIMVAL series was to review current efforts in simulation validation and support technical interchange on simulation validation (SIMVAL). Since it was apparent at the time that much of the difficulty in communication regarding M&S — both within and across the different organizations and Services — was simply due to the lack of common terminology, another purpose was to arrive at a common set of definitions. As one of the products of SIMVAL, it was desired to develop a plan for future efforts to address issues on simulation validation.

October 1990 MORS SIMVAL Workshop

The initial MORS SIMVAL workshop was held in Albuquerque, New Mexico in October 1990.[2] It looked at several aspects of M&S VV&A, including requirements analysis, system design, Operational Test and Evaluation (OT&E), operations support/tactics development and training. Separate sessions were held addressing each of these applications to highlight the similarities and differences in VV&A practices.

One of the concepts discussed at the initial workshop was a "level of validation." Validation is a process, and the adequacy of any validation is a function of the intended application of a particular model or simulation. It would be nice if levels of validation could be defined such that one could require that the level of validation be a function of the complexity of the issue being addressed and the importance of the analysis being supported. A model being used in an acquisition decision, for example, would require a higher level of validation than one used to help plan a test. Levels could be associated with factors such as configuration management, logical verification, documentation, code verification, face validation, independent review, comparison with lab data, comparison with operational data, etc. While the concept was rejected on a practical basis, it did convey the concept that validation is a continuing process focused around specific applications of the model being used, with accreditation being a decision that the degree and type of validation is adequate for the intended purpose.

An assumption of the workshop was that there are several "common-use" models – models with several users with the same or slightly different applications. This idea was addressed in an earlier MORS workshop, "ADMAS".[3] That workshop looked at several models with different applications in order to identify models with high potential for re-use. The approach used was to define the application, determine the elements of the model that were important to that application, and then analyze the degree of fidelity necessary for that application. That approach is directly applicable to model V&V.

One of the problems discussed at the workshop was the cost of performing V&V. While the cost of ensuring that a model is appropriate for a given application requires time and effort, if there are several users, the verification and validation databases should have shared benefits. Validation for the initial purpose – the original intended application – would be tailored for that application. However, subsequent applications could take advantage of that validation; adding on in areas where the application required a different emphasis. This increasing validation database should result in cheaper V&V. This would be especially true in requirements for areas such as target signatures, terrain effects, etc. – where common requirements would be likely.

For system acquisition, it was also assumed that models used in design, development and testing should become a normal part of the contractors' deliverables. Common areas – like terrain effects necessary for each model use - could be shared. It was recognized that model V&V and use within an application could be expensive, but model use across applications should be efficient.

There was considerable discussion at the workshop regarding the lack of V&V in many applications, and frustrations expressed from the operations research community that many of the model and study customers were more concerned about the models providing answers that coincided with the desired outcome rather than one which provided a true insight into the issue being addressed.[4] V&V was seen as a way to ensure that the model provided a source of objective analysis rather than as a public relations tool.

The workshop attendees concluded that:
- It is feasible and desirable to define methodology in model validation;
- It is not reasonable to specify a standard scale or a numerical rating for a validation method;
- Model validation should be a tractable process that is repeatable;
- Model documentation is critical, as is documentation of any validation efforts and steps taken leading to accreditation;
- Accreditation must be with respect to a specific application;
- Data for model validation must be available and validated for the same range of uses of the model;
- Accreditation should be accomplished by someone who understands the specific problem to be addressed and has responsibility to address that problem.

10

Definitions derived at the workshop were:
- *Verification*: The process of determining that a model implementation accurately represents the developer's conceptual description and specifications.
- *Validation*: The process of determining the degree to which a model is an accurate representation of the real world from the perspectives of the intended users of the model.
- *Accreditation*: The official determination that a computer model is acceptable for a specific purpose.

These definitions were briefed to the MORS Sponsors, and subsequently adopted as the official VV&A definitions for use in DoD policy. Subsequent SIMVAL workshops addressed the elements of VV&A, as well as other VV&A topics.[5]

February 2001 MORS SIMVAL Workshop

The report from the 2001 Reno, Nevada MORS workshop titled "Assuring M&S Credibility for Defense Acquisition and T&E: Survivability, Lethality and Mission Effectiveness" concluded that the OT&E community has an M&S and VV&A policy and that it seems to be more rigorous about evaluation of M&S than other communities.[6] The Reno workshop also observed that it is not unusual for the OT&E community to reject the use of M&S due to lack of data to support validity, and that the OT&E community's expectations of M&S outstrip current status of, and progress in, M&S validity.

The Reno workshop concluded that there are significant advances in VV&A science and technology that need to be made, particularly in the areas of correlation and sensitivity analysis. In addition, there needs to be a stronger emphasis on the mathematical underpinnings of validation and accreditation to allow for movement from qualitative to quantitative assessments. "The foundations for credible M&S used in support of the acquisition process would benefit from a more disciplined approach to M&S development, assessment (to include VV&A) and implementation."

October 2002 MORS SIMVAL Workshop

At the MORS SIMVAL workshop held in October 2002, one of the observations was that "the central problem is who uses them [models] and how they are used. VV&A has been a dead weight on the use of M&S." This a rather harsh view of the progress made in model VV&A, but one that is held by many in the community. To quote Dr. Ernest Seglie and Dr. Patricia Sanders, "The formal process of accreditation should mirror the processes that accompany all research: the process of convincing oneself that the methods and results are reasonable, appropriate and worth believing to some extent, and of assessing the confidence one has in the results."[7] From the perspectives of some attendees, VV&A has focused too much on the bureaucratic aspects of the problem, with the technical requirements addressed qualitatively through SMEs rather than through objective analysis. While subject matter experts will always play an important role in validation, the primary means should be comparison with real world data. When we can

learn to focus on the real issues of M&S, VV&A should become that natural process of ensuring that a model provides an answer adequate to address a given problem.

The MORS SIMVAL series has revealed that VV&A of models and simulations, in and of itself, is not the problem in using M&S to support T&E. Rather, there are larger, more fundamental issues that must be addressed before the proper role of M&S in T&E can be determined. One of the products of the October 2002 MORS workshop should be an identification of these issues, as well as an action plan that will move M&S to a more efficient means of testing systems to support the warfighter.

Endnotes

1. "A Base of Sand Problem: A White Paper on the State of Military Combat Modeling," RAND Note N-3148-OSD/DARPA, by Paul K. Davis and Donald Blumenthal
2. MORS "Simulation Validation Mini-Symposium (SIMVAL)," the BDM Corporation, 15-18 October 1990.
3. For a discussion of using common models for different applications, see MORS Workshop Report "Air Defense Models and Simulation (ADMAS)," July 1987.
4. Also see "Report Of The Defense Science Board Task Force On Advanced Modeling And Simulation For Analyzing Combat Concepts In The 21st Century," May 1999.
5. Also see the following MORS reports: "SIMVAL II," The Institute for Defense Analyses, 31 March-2 April 1992, "Complexity in Models and Simulation Mini-Symposium," BDM Albuquerque, 24-27 February 1997, and "SIMVAL '99: Making VV&A Effective and Affordable" Johns Hopkins University Applied Physics Lab, 26-29 Jan 1999.
6. "Assuring M&S Credibility for Defense Acquisition and T&E: Survivability, Lethality and Mission Effectiveness," Reno, Nevada, 12-15 February, 2001
7. MORS SIMVAL II (ibid)

Modeling and Simulation: *Is VV&A the real stumbling block? Are we using M&S correctly? What is V&V if we use models correctly?*
Dr. Ernest Seglie

The National Research Council's Academy Industry Program sponsored a two-day symposium on Mathematical Modeling a while ago. This was a meeting of committed enthusiasts who, at least on the industry side, acknowledged that modeling has not worked as often as they would like. The subjects of general interest most often mentioned concerned:

- The **context** in which the modeling work is to be done.
- The **best uses** of modeling.
- The **environment needed** for successful modeling.
- A discussion of **reasons for the lack of success** in use of models.

The reason this is important to a conference on VV&A is that if we are asking the wrong questions, or have the wrong expectations, of models, then we may never be in a position to validate them. This is very important, especially in testing, which will be the primary example used in this discussion.

Context for Modeling

1. Models should *help* critical thinking; then go test.
2. Modeling is the hypothesis step of a scientific method. It generates the "I'll bet 'ya'" type statements that then should be tested.
3. Models provide insight they do not make decisions. They are not an answer-machine. "Good judgment comes from experience and experience comes from bad judgment." There is an approach in which decision makers get experience with the models – where the losses are only 'virtual.'
4. Models provide a framework around which quantitative questions can be asked.

Best Uses of Models

1. Hypothesis generation.
2. Sensitivity Analysis – models provide the logical consequences of explicit assumptions.
3. Generation of imaginable situations. The "outlier" results are important and the reasons should be clearly understood. The range of situations and the range of responses are important to know. In general, policy makers need to appreciate the outliers as well as the most probable outcomes.
4. Scenario generation. "Unpack" the distribution to look at the 3-sigma case. The danger in scenario generation is that the assumptions become hidden.
5. To facilitate a "rolling reassessment," adaptive management capability. This is as part of a feedback loop of Operations -> Field Data -> Data Analysis -> Quality Controlled Database -> What-if analysis with models -> Operations. This looping is hard to start but becomes self-re-enforcing with each loop.

6. There are now tools to explicitly model the uncertainties in a problem (such as diagnosis in emergency rooms) and suggest further testing to reduce uncertainties.
7. When, as in some cases, there is nothing else to do. (Sandia-nuclear case)
8. In some cases, nothing else will do better. Mostly to crystallize thinking on some point.
9. To accelerate knowledge. To help in the visualization of results or technology transfer. The workers in the program must internalize the results.
10. Sharpen critical thinking.
11. To suggest experiments that are difficult, and would not get done if the importance of them were not demonstrated by modeling.

The uses advocated most consistently are very robust. There is little need for a discussion of validation and accreditation, or extrapolation. The uses endorsed almost make the models "self-correcting," at least in the sense that the team is actively looking and testing for where the model might begin to break down. There is never a suggestion that modeling replace testing, but rather that modeling could, by helping to illuminate the areas of greatest uncertainty, guide the allocation of testing resource to accelerate learning. The one validation comment, by the climatologist, was to be sure to validate at the same resolution that the model uses. Models of a single cloud may appear quite good, but to use them over an area the size of Connecticut might be bad because there could be twenty such clouds in that large area.

Environment Needed for Successful Modeling

1. Multidisciplinary team including operational expertise, engineering, physical understanding, mathematical, statistical, etc. This item and the suggested use of models as a visualization tool for technology transfer work together to make modeling a way to get the most out of an interdisciplinary team.
2. Understanding of the fundamental processes modeled.
3. Mathematical understanding. Models must be analyzed and understood well enough to reformulate them when required.
4. Computational power; though this becomes a crutch if used too often.
5. The use must be timely and meet other resource constraints
6. Ease of use determines use. From the industry perspective, spread sheet models were the ones used most, because people understood them. This is true even when they generate deterministic and point estimates. If greater detail or more information is warranted, spreadsheet combined with probability models are used to generate risk analysis and probability distributions. Finally, operations simulation with dynamic modeling is used when warranted to take account of event uncertainties, and multiple scenarios.
7. Models must account for people and this is difficult.
8. To make the modeling activity work requires the right data at the right time and at the right place.
9. Integration of diverse models is difficult and requires a special environment.
10. Simulation has to have feedback from field data.

Lack of Success In Use of Models

There have been studies of the reasons why the promise of models has not been realized. Chief items for considerations are:
- Inadequate physical modeling.
- Inappropriate method of formulation or representation of the problem.
- Inefficient computational procedures or the use of a COTS software environment.
- Ill-conceived multi-disciplinary integration.
- Lack of procedures to assess magnitude, propagation and impact of uncertainty.

The explicit identification of uncertainty needs to be clear to the audience. At the Academy forum, someone had been working on a taxonomy of uncertainty. The list of how his uncertainty could arise included:
- Missing components or potential errors in inputs.
- Known processes but unknown functional relationships.
- Known structures but unknown or erroneous values in some important parameter.
- Unknown appropriateness of underlying assumptions.
- Noise in data, bias or incomplete observation.
- Use of historical data that does not account for changes over time.
- Uncertainty arising from stochastic nature of process.
- Uncertainty due to projections.
- Human behavior.
- Numerical processes.

In addition to the above reasons why a modeling effort can fail, we can add:
- The divisive effect of jargon to the communication between multidisciplinary team members.
- Few existing value/reward systems actually reward collaboration.
- Funding sources deal awkwardly with multidisciplinary work.
- Isolation of the modeling community.
- Difficulty in integrating models of different resolution. Integration of diverse models is difficult and requires a special environment. Integration often leads to the wrong answer if the wrong resolution is used or if the scale of resolution of the model is not the same scale of resolution of the observation.
- The practical sociology between modelers and decision makers

Conclusion and Implications For Testing

Despite the fact that the Academy forum participants formed a pro-modeling group, the uses suggested and the cautions implied are quite sobering. If anything, disappointment was expressed by those who had presented model results to policy makers. They felt that policy makers at high levels were just interested in an answer such as: what will happen. Instead they should be interested in what can happen, what are the odds and how do we know. Some forum participants felt that the context in which "model" answers are

15

generated are important and not often explained. In short, modeling should be an aid to clear thinking not a substitute for it.

How could we in the DoD test community incorporate these insights while not losing the potential benefits of modeling?

I believe we should find better ways to encourage the programs to integrate modeling into their development and decision process. With respect to testing, the OTAs should not be forced to lead in this process. They should not propose the use of models and simulation as part of their IOT&E until the modeling has demonstrated it is an integrated part of the processes in the program and an aid to clear thinking. The conditions mentioned would roughly translate in the following way. The proposal to use modeling should wait until:
- A multi-disciplinary team is established (engineers - both government and contractor, operators and testers), in place, and functioning to add information to contractors, program offices and evaluators. This capitalizes on one of the chief named benefits of modeling, viz., as a communication tool, and has to be done early in the program- seven to ten years before the IOT&E.
- The feedback loop mentioned in "Best Uses #5" has worked through the cycle on early testing of components and sub-components, and the model has been used in some early operational assessment.
- The model has been developed in sufficient resolution so that the testing planning parameters (for example the scenarios to test, the number of replications, and the variability of output measures) can be, and are, calculated using the model.

At that point the model should be able to provide on-going evidence that it is a useful tool to aid clear thinking, is integrated into the program processes, and is contributing to the program. This evidence should include its assisting early testing in hypothesis generation, scenario generation, sensitivity analysis (as components mature to final performance values), identifying important information voids to be filled for tests, and to suggest experiments, and design modifications or operational concept modifications. All this is important because the OTAs cannot independently support the modeling effort. Model use that is OT&E centered is not the goal. The modeling has to have wider support in the contractor and program office. The utility of the modeling should be to the program as a whole, not just to the OT&E part of the program.

In the TEMP preparation process, the modeling should provide estimates of the resources and scenarios needed in the IOT&E (for example, number of test items, number and type of scenarios, other resources such as threat type and size). In other words, the modeling should be able to demonstrate its utility as a tool for clear thinking in the on-going program and the test planning process. A specific case of such use might be in the determination of the number of LRIP items needed for the test. The determination of that number could require the presentation of model results that included the variability of system performance and performance across scenarios.

As part of the test concept brief to OSD, the model, using the current values of system parameters from testing to date, should predict the OT&E results and provide sensitivity

analysis around those points. In particular the model should be used to estimate the probability of the system passing OT&E. The model should be used to help determine which tests should be run first to gain the most information as early as possible. (This is a different approach from the public relations approach taken by most program offices of doing the easy tests first to get successes under their belt before trying more difficult and realistic tests.)

The above suggestions amount to a shift in emphasis. I believe the Department would be better off if models were integrated into the program as a whole in the ways suggested. Our efforts should be directed to that. Programs that proceed until the end without models and then use models in the evaluation are not getting their full benefit. Evaluations designed to depend on models of a system that do not have a long history of useful feedback increase the risk of an inadequate evaluation. Therefore, what effort we expend encouraging the use of models should be to encourage, indeed require, their use early and in an integrated fashion, or not at all, in the evaluation.

M&S and VV&A in OT&E
Dr. Marion L. William, FS, Chief Scientist, AFOTEC
Mr. Brian Barr, Technical Director, ATEC
Mr. Steven Whitehead, Technical Director, OPTEVFOR

When the original MORS SIMVAL workshop was held in 1990, there were some implicit assumptions about how DoD would treat the use of M&S in the acquisition process, and how those uses would support M&S use in T&E. We – the testers – generally assumed that models would be developed as a natural part of a system's acquisition, and that we would simply take advantage of those efforts. Validation and verification became a key to our use: We would use the V&V from the "original" application, and add on any specific V&V needed to account for differences in applications.

Among the specific assumptions were:
- **Common use models**. A model developed by a contractor for a specific weapon system as a part of the design process would be used by development testers as well as operational testers. In an ideal situation, models used in the Analysis of Alternatives (AoAs) would form the upper level (campaign level) part of the M&S spectrum. Such models would help identify the driving factors in system employment, and assist in conducting sensitivity studies. Engagement and engineering level models would provide the foundation for the upper level models and would provide much of the M&S support for T&E. With multiple users, there would be a better chance of finding model errors. As model use moved from design and development into test and evaluation, the different model applications would exercise the model in different areas, and would allow several opportunities for comparison of test results and model results with different sources of data. With time, these models should provide an increasingly accurate representation of the real world — or at least the parts of the real world that were important in the applications being addressed.
- **Improved databases**. One of the hassles in model use is getting valid data bases — including such areas as terrain and target signatures. A systematic and structured process for model use, as expected from the increased attention to M&S, would result in better and more and readily accessible databases. In a recent experience, finding a bi-static radar signature involved not searching existing databases but calling around the country until the right person was contacted who knew of the existence and basis for the needed signature data. Despite the emphasis and attention on M&S over the past ten years, M&S use hasn't improved to the degree anticipated.
- **Identification of common requirements**. There are several areas where improvements could be made in existing databases. There are numerous requirements and uses in common areas such as terrain (multipath, clutter), signatures (IR, RF), etc. Gathering common requirements and creating a cooperative network to share existing data and data validation would significantly improve the efficiency and creditability of model use. It would seem feasible to

have organizations with the responsibility of overseeing model data and data validation - such as MIT Lincoln Laboratory for terrain data.

In general, DoD and Service policy considers operational tests to be only those tests that involve the "real" system. The preference is always a real system in a real environment, but often test and resource limitations prevent a true operational scenario. In these cases, the real system – or parts of the real system – is subjected to a simulated environment. The use of hardware-in-the loop facilities is used to simulate a real hardware system with real signals produced by an artificial environment. Whether this is adequate for OT&E depends entirely on the question being addressed. Rarely is a model of the system within the model of the environment acceptable as "OT&E data."

While M&S has been used in OT&E for quite some time, there has been an increasing emphasis on it over the past several years. However, in many users' experience, the results haven't matched the expectations. Consequently, the operational testing community has revised its expectations for M&S. Consensus seems to be forming around a more limited notion of the proper role and function of models in OT&E. There are certain uses for which models can be particularly useful and ought to be used, and others where model use should be limited or not used at all. Models cannot replace operational testing but can be valuable tools, provided they are verified, validated, — and above all, accredited — for their intended use.

Areas where models have great potential to support OT&E are as a pre-test planning tool, as a pre-test prediction tool, or as a post-test analysis tool. Models also hold promise, as tools for gaining insight into issues of operational effectiveness or suitability that are unlikely to be discovered through conventional operational testing.

In pre-test planning models can help identify areas where testing should be most profitable; perhaps not reducing the amount of testing required, but at least identifying areas with high risk, or areas with the highest payoff for testing. Since this enables operational testers to focus limited test resources and articles on higher-risk areas and reduce testing in lower-risk areas, this is the most cost-effective use of M&S in OT&E. For pre-test prediction, M&S can help establish data and instrumentation requirements.

In cases where M&S was used as a planning or prediction tool in Air Force OT&E, there were often areas where a model provided insight that helped plan a test, or identified areas where instrumentation was required to better understand the test results.

There have been a number of good examples where M&S has provided insight into system performance that could not be obtained, at least not affordably, in an operational test environment. Several years ago it was discovered that some electronic countermeasure techniques were sensitive to normal manufacturing tolerances. Many countermeasures were designed to affect specific circuits within a threat system, therefore depended on knowledge or assumed knowledge of the design of that system — sometimes using comparable technology. However, countermeasures that were effective against one threat serial number were not effective against another threat of the same type

but different serial number. Normal manufacturing tolerances make the circuits within a production run different enough that countermeasures were effective only a portion of the time. Testing against a large sample of missiles was not possible due to cost. However, it was feasible to develop a model of the threat system, validate the model against hardware test results, and then vary the model in the same way the hardware model would vary under manufacturing tolerances. The model gave a valid insight into which countermeasures were effective across the range of manufacturing tolerances — and across the ranges of missiles that could be encountered in combat.

M&S use in Operational Assessments (OA) — where realistic operational testing has not been accomplished but where there is sufficient design and development information available for the testers to at least identify the high-risk areas in an acquisition program — has been especially effective in programs where M&S is a normal part of the design and development process. Air-to-air missile programs such as AMRAAM and AIM-9X are prime examples. In these programs, the contractors and program offices have a structured program to V&V the models, with continuous feedback from test data to the model to correct any coding errors and to improve the model.

Unfortunately, models are not typically delivered as part of the acquisition process of DoD systems. While models are often developed by the contractor for internal use, they are not available for OT&E and/or are not documented — primarily due to the fact than no one paid the contractor to provide a model as a deliverable. Where models are used in trade-off studies, the study results, including assumptions, are rarely available — again because it was not a deliverable on the contract.

So while the Operational Test Agencies (OTAs) are a major "consumer" or user of M&S, they are usually not the developer. There are exceptions where they develop simulations specifically to support T&E or enhance simulations developed by others. But on the whole, attempts to develop or make major modifications to models by operational testers to adapt them for their own test applications invariably take longer and cost more than estimated. As a result, M&S value in T&E – at least in OT&E – has been limited.

The notion that the development, verification, and validation of models be left to industry and be contractually tied to programs has been suggested. As part of the acquisition process, this notion holds, industry should be required to present their M&S plans when responding to RFPs and cost out that effort. By letting industry take the lead in model development, they would use their own business models and processes to decide where M&S development is profitable. If they determine that it is not profitable, they wouldn't do it. But if they determine that it is, they'd eagerly do it, passing the cost, along with the models, on to the consumer.

Early efforts to get Service endorsement of the use of M&S as a normal part of the acquisition process have produced mixed results. One issue has always been the lack of M&S requirements in the Program Management Document, and funding for model development and VV&A. Where M&S has been advertised as a means of increasing the efficiency of testing, the funding for testing has sometimes been reduced without

providing the funding for model development. As models are developed and used for different applications within an acquisition program, and used by different acquisition programs, the efficiency and effectiveness of M&S will increase. M&S is cost effective across systems, but not always within a single system's acquisition program.

Another issue in making model development and VV&A part of the acquisition process is potential incompatibility with the new DoD acquisition model. In moving away from soliciting industry through an RFP process based on needs and requirements and inviting industry to come forward with ideas for delivering needed capabilities, it becomes difficult to make models a requirement. If industry happened to develop models in creating their ideas for delivering these capabilities, the models might be available for use in T&E. But if industry had not developed any models, it would be too late to go back and develop them after they had presented the ideas to the DoD for consideration.

Regardless of how models are developed, verified and validated, any model used in OT&E must be accredited for its intended use. Since the OTAs are usually not the simulation developer they are also not responsible for V&V. That is not to say that they are indifferent to V&V. Quite to the contrary, the OT community seems to be more thorough when examining the documentation and V&V of models than other communities. It is not unusual for an OTA to reject a model due to V&V of dubious rigor or credibility. Most OTA policy is actually directed towards accreditation. Unlike V&V, accreditation is the responsibility of the model user (in this case the OTA). Accreditation is the official determination that a model or simulation is appropriate for the intended use.

Within ATEC, there is a formal process for accreditation. It starts with an accreditation plan that must be approved by the accrediting authority. For major systems, this is the ATEC Commander (usually delegated to the Technical Director.) The heart of the accreditation plan is selecting appropriate criteria upon which to measure the acceptability of the model. After the accreditation plan has been executed, a formal report is presented to the accreditation authority for his approval.

The biggest dilemma faced in this accreditation process is the requirement for test data to accredit the model. Often those data are not available until late in DT or sometimes early in the OT. This leads to a very short timeline for completing the accreditation report and no way out if the accreditation is disapproved.

It is important to keep in mind that accreditation is always use specific. Just because a model has been verified and validated to support one test, it does not mean that it has been or will be accredited to support another test.

Models will improve as testing is accomplished and test results are compared with model results. Although this is an advertised requirement in DoD policy, it is rarely accomplished; at least not on a routine basis. Dozens of operational tests are conducted each year by military organizations. In most of these tests there are models corresponding to the test scenario and test systems that should provide a valuable opportunity to collect test data for correlation with model data. However, this is the

22

exception rather than the rule. Even in examples where scarce resources are expended in live firing, getting the right data for comparison with a corresponding model isn't accomplished, as it should be.

There are several valid reasons why this occurs: Every one agrees it is a good idea to improve models with real data, but an organization has to focus on getting time and funds to accomplish its assigned tasks. Getting data for "other people's models" is not one of those assigned tasks. There have been a number of instances where old test data was investigated as a source of data for model validation. This is almost always a fruitless effort – planning for data capture for comparison with model results requires effort before the test to ensure that the right data, along with the right accuracy, are captured.

Fundamental problems include the fact that there is no central DoD-level—or even Service level—agency acknowledged by all as model and data base keepers or archives. Indeed, there is currently no structured process for model verification and validation. Such an agency might take on the additional task of identifying "validation data sets" of data obtained from real-world tests to be used in model refinement and validation.

In conclusion, M&S in OT&E still holds great promise of increasing the efficiency of OT&E. However, the current process must be changed to make it happen. Some of the general changes that will be required include:

- Early, integrated test concept along with M&S requirements, and with a model of a system as a contract deliverable.
- "Surgically-defined questions to be address by the model." No model can reasonably represent the real world. A model can, however, represent those parts of the real world that are important in the application being considered. Trying to build any model without precisely defining the question to be addressed is a fundamental, but not uncommon, mistake. And it is often fatal.
- Along with defining the question to be addressed is the definition of those parts of the model that are important to that question. This is the fundamental step in model validation. The results should be a well-constructed V&V plan to ensure that the important parts of the model are adequately represented. This can only be accomplished with a pragmatic V&V plan with joint execution by those agencies that have interests in the model – design, development test and operational testing, as an example.
- For the effective use of M&S in T&E, we need to move from the current ad hoc approach to a more structured and disciplined approach. This is unlikely to occur until there is a strategic plan with objectives that can be implemented to start us moving in that direction.
- While significant advances in VV&A science and technology have been made, more are needed, particularly in the areas of correlation and sensitivity analysis.
- There must be a stronger emphasis on objective quantitative assessment of models rather than assessment by SMEs.

MORS

Working Group 1 Outbrief
Use of T&E Data To Support M&S VV&A

Test and Evaluation, Modeling and Simulation and
VV&A: Quantifying the Relationship Between Testing
and Simulation

MORS Workshop
Kirtland AFB
15-17 October 2002

Contents

- Introduction
  - Members
  - Objectives
  - DOT&E Letter
  - Approach
- Background
- T&E Data Considerations
  - Availability
  - Structure
  - Culture
- Recommendations

Introduction

Members

MORS

- Mr. Fred Hartman, FS   JSIMS (Co-Chair)
- MAJ Todd McDonald   HQ AFOTEC/CNR (Co-Chair)
- Members

Raoul Anaya	Dr. William Baker	Tommy Blair
George Cheng	George DeGovanni	Robert Easterling
August Fucci	Michael Goodwin	Dr. Anne Hillegas
Capt Jesse Kemp	Ms Michelle Kilikauskas	Ken Konwin
MAJ Doug Matty	John Illgen	Ms Lana McGlynn
Jack Sheehan	Leonard Vance	Gery VanderVliet
Ms Diane Verner	Charles Walters	Dr. Steve Warner
Luis Castillo	Ronald Arlund	John Walstrum
John Steele		

Introduction
Working Group Objectives

- Support the larger Workshop Goals and Objectives
- Explore guidance and develop recommendations based on the experience of the M&S community in the use of Test and Evaluation data to support the VV&A of M&S
- Highlight potential ways of applying data to foster reuse and improvements in T&E

Pre-test/Post-test

Make a distinction in data. Completing the feedback loop: requirements for simulation support plans.

Holistic view of data: how does data flow back? Taxonomy of how data fits into overall process (ex: MDA – not sharing data based on $$).

Sharing of data and who controls it. Contractor requirements to populate an existing database; Clout at administrative level to make it happen; Government organization oversees.

Impact of DMSO/MSIAC facilities - Authoritative data

Define a broad structure of what the data needs to look like. Know and define parameters; sample size; sampling interval. Needs to be stated. Point estimates or statistical analysis that is inadequate but, no cookbook or template

Repository: needs to be distributed

Each program needs to maintain its own database. Is it worthwhile even to have? We've had them for ten years. The reason they are not populated is cultural, it is not a technological issue. Capturing metadata about a flight test. Capturing metadata about VV&A. "IACS" as a repository?

OSD policy briefing - Thoughts on data management. Scope too narrow... Quantitative aspects are fairly easy to identify. Validated models may not have reference of original data sources. Precision- tradeoff and analysis- what do you need. Does your data need correspond to the actual need (Ex: earth penetrators, 1 million elements.. Choosing the right Aspects to model..)? How much is enough? Are resources commensurate with results? Getting threat data is easy compared to getting data on blue systems. SPO... approval: contractors can't get it, government can't get in many cases.

Introduction

m⊕RS

Points from DOT&E Letter on M&S to OTAs - Mr. Christie, Jun 02

- Test as we fight:
  - Real <u>warriors</u> employing real <u>combat systems</u> conducting realistic <u>missions and tasks</u> in a representative <u>physical environment</u>
- Core T&E processes:
  - Prediction (hypothesis), planning (test design), data collection (test event), analysis (data verification), reporting (conclusion)
- M&S supports — but does not replace testing:
  - Before (sizing, scoping, design), during (stimulation, response), after (interpretation, significance, assessment)
- Holistic:
  - Constructive participation and engagement of key stakeholders through the iterative model-test-simulate-evaluate cycle

Introduction
Working Group Approach

mors

- Explore the use of T&E data to support M&S VV&A
  - Policy and Procedures
    - Primarily covered by other WG 2-3
    - "Basis of Confidence" issues
    - Implementation difficulties
  - Experiences and lessons learned since SIMVAL 90
    - Reuse: data quality uncertain and not traceable
    - Availability: tenuous
    - Program planning: assumes presence of data and proceeds
    - Examples: DMSO-MSRR,OSD PA&E-JDS,DIA-TMAP

Very few programs have data investment strategy.

You need to have a senior person advocating data reuse and sharing.

There is usually no money left after OT&E to fold data back into any repository.

No mechanism exists for folding data back in, even if you wanted to.

"We need a way to keep track of data - where it is at: DTIC like"

But it already exists: DMSO has authoritative data source project that has existed for 5 years

Must be linked to AT&L at some point... larger picture of acquisition first priority and focus of T&E community though.

Introduction
Working Group Approach

- Explore the use of T&E data to support M&S VV&A
  - Limitations to acquiring and sharing data
    - Resource requirements – collection and maintenance
    - Not there, not enough, no access, no context/pedigree
    - Lost and forgotten data: collected, stored, and never used
  - High level advocate is required
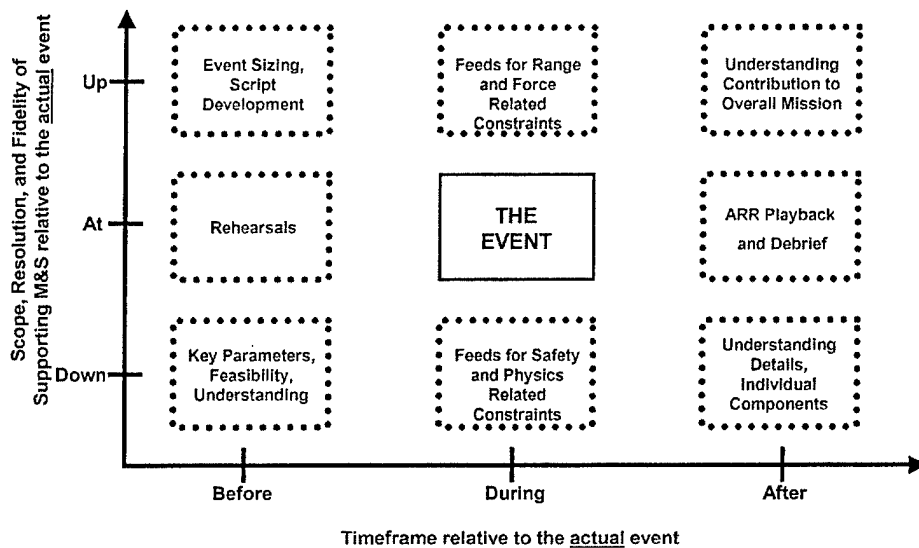   - Need executable plans

BACKGROUND

Background
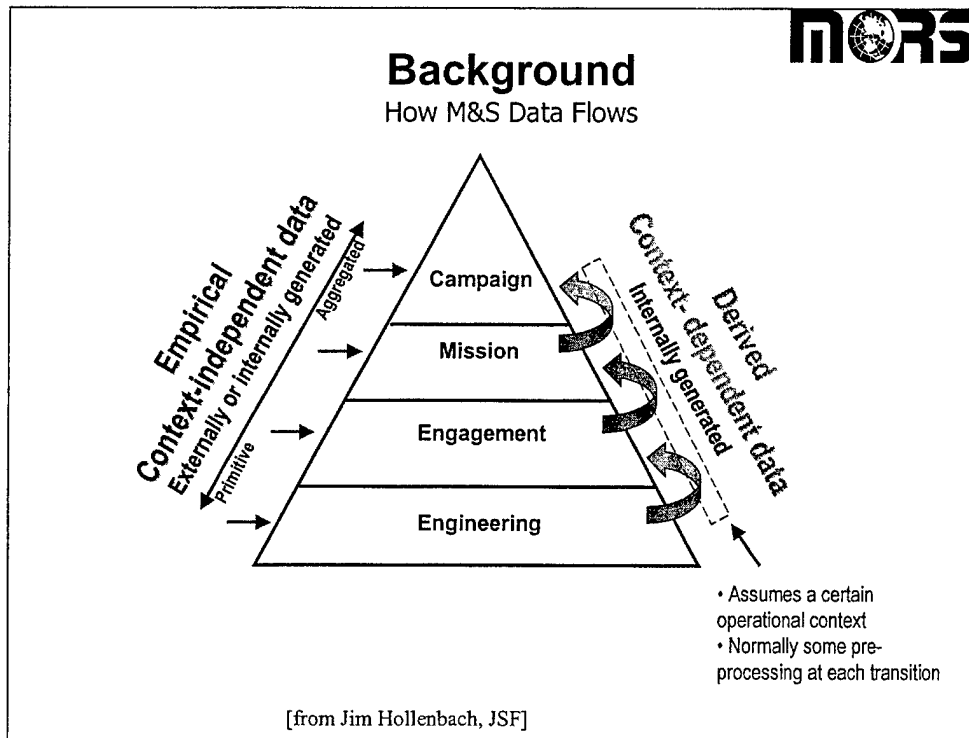
Factors Pressing for more M&S in Test and Evaluation

- Substantial increase in operational mission complexity and required integration of complex system of systems
- External encroachment on existing ranges in the face of significant increases in geographic stand-off between new and emerging operational systems and targets
- Desire to simultaneously reduce the time-to-field new systems and reduce life-cycle systems costs
- Desire to simultaneously increase T&E realism while reducing M&S costs
- Technology push from the modeling and simulation community

Background
Integrated M&S, T&E Support for Knowledge Formation

Scope, Resolution, and Fidelity of Supporting M&S relative to the actual event

Up
- Event Sizing, Script Development
- Feeds for Range and Force Related Constraints
- Understanding Contribution to Overall Mission

At
- Rehearsals
- THE EVENT
- ARR Playback and Debrief

Down
- Key Parameters, Feasibility, Understanding
- Feeds for Safety and Physics Related Constraints
- Understanding Details, Individual Components

Before | During | After

Timeframe relative to the actual event

Background
How M&S Data Flows

Campaign

Mission

Engagement

Engineering

Empirical
Context-independent data
Externally or internally generated
Aggregated
Primitive

Derived
Context-dependent data
Internally generated

• Assumes a certain operational context
• Normally some pre-processing at each transition

[from Jim Hollenbach, JSF]

Measured data vs derived data

35

Background

- To Make a Difference, M&S and Data Should Be:
  - Credible: Make hard decisions and have them stay made
  - Timely: Cycle times of hours and days, not weeks and months
  - Affordable: Faster, easier and cheaper than alternatives

Credible: Which of the possible event scenarios provides the most important information? Do the actual event conditions achieved provide reliable operationally-realistic parameter estimates? For the observed performance, what is the operational effectiveness, survivability and suitability of the forces and materiel under evaluation?

Timely: Need M&S setup times that are commensurate with execution times for pre-event predictions, event execution and post-event evaluation.

Affordable: Existing development, setup, verification, validation are complex, labor-intensive activities (but it doesn't have to be)

T&E DATA CONSIDERATIONS

T&E Data Considerations
Topics

- Is T&E data **available** from T&E community (DT, OT, threat assessment)?
- Does T&E community support a **structured** data program?
- Is a **culture** change needed?

T&E Data Considerations **m⊛RS**

Availability

Is T&E data available from T&E community (DT, OT, threat assessment)?

*   Who owns the data?: USAF, USN, USMC, USA-PM/OTA, DT agencies, intelligence centers
*   Why don't owners make their data available?
    * $$$$ (won't collect data for other purposes)
    *   Quantity of data
    *   Unknown user list
    *   Policy issues (Title 10)
    *   Source selection
    *   Data certification

T&E Data Considerations

Structure - Support

Does the T&E community support a structured data program?

1. Taxonomy
2. Accessibility
3. Business Case

Look at prior studies that attempted to define "Authoritative Data Sources:" DMSO; goal was high level view of data at beginning, but went well into details.

Look at other organizations (American Institute of Engineering and Astronautics): see how we can leverage or use how other organizations manage data taxonomies.

What is the buy-in cost of T&E structured data program?

T&E Data Considerations

Structure - Taxonomy

- Taxonomy considerations
  - Data repositories must be useable:
    - Need formats, standards, recognizable content
    - Need metadata
    - Need technical structure for accessibility
  - You must STORE as part of the metadata the CONTEXT of the test data itself (pedigree)
- MUST associate the circumstances and test environment that is associated with test data
- The limitations of test data in a repository are not apparent
  - Qualify HOW data was collected (uncertainty, quality)

T&E Data Considerations

m⊕RS

Structure - Taxonomy

- Would a test data taxonomy be beneficial to the community?
  - Data must be found prior to inclusion in repository
  - Data must be in recognizable formats (standards)
  - Data must be accessible
- Taxonomy considerations
  - Example: Consistent/common threat data may be more important than common validation repository
  - We need to have a clear definition of data: point estimate data stream, ~flyout models
  - Who is the user and what is their purpose for the data?
  - There is a vast difference between data requirements for a user of a simulation than the data requirements for the VV&A or IV&V of a simulation
  - Who determines what data is appropriate?

Do we even need high level of detail for our own blue systems if we are comparing them against red systems with far less resolution?

T&E Data Considerations

Structure - Accessibility

- There is an investment currently lost (shoebox approach)
- MOAs and Beg/Borrow/Steal have to be used in most cases to get to data; PMs are the guard post that cannot be by-passed in most cases
- Classification issues, sensitivity, and proprietary nature of data restricts MOST of the access to it
- PEOs controlling PMs may be the best source of Mandating Data control (MDA)

T&E Data Considerations
Structure - Accessibility

A data repository is like a drawer full of sharp knives:

Adult supervision is required...

T&E Data Considerations
Structure – Business Case

- Does the value of having a repository outweigh the price it costs to adequately populate it?
- Will there be data managers that data passes through (trusted gatekeepers)?
  - How/who will be responsible for the repository?
  - How/who will be responsible for entering the data in the repository?
- Data should be posted according to appropriate security, proprietary, source selection (contractual)
- There should be a centralized infrastructure that hold metadata and points to location of data
  - Build metadata as you build the data

Will the data manager be an impediment to the process?

T&E Data Considerations

Structure – Business Case

- Establishing data in a repository is one thing: updating the data (changing it for new needs) is a different and more important issue
- If it is not a deliverable in the contract, you won't get it!
- Need incentive structure for sharing of data
- COST: $$$$$ to populate the database
- SCHEDULE constraints of programs to generate data

Due to proprietary/source selection issues, a lot of data is just not available.

T&E Data Considerations **MORS**

Culture Change – Data Ownership

- "Ownership of Data Issue"
  - Proprietary data
  - Contractual considerations
  - Costs to populate and maintain
  - Costs to document for reuse
    - Contextual limitations of T&E data not apparent
- PMs need permission to share
  - Model validation = need to know
- Transformation MUST be accomplished in the way that data is viewed
  - Use MORS as a forum to educate community on WHY you need data sharing
  - Build a path to document within OSD standard formats

T&E Data Considerations

Culture Change – Data Ownership

Limitations are more cultural/business than technical

T&E Data Considerations   ⅢⓈRS

Culture Change – Technology

- Existing technologies for centralized/ distributed data management will support vast amounts of data
  - Metadata is order of magnitude greater than the data
- Including context with data enables data mining technology

WG 1 Recommendations

1. Make T&E data more available for M&S applications
2. Support a structured data program
   - Taxonomy
   - Accessibility
   - Business Case
   - Culture

WG 1 Recommendations

3. Identify potential control centers for T&E data
   - OT&E: DOT&E
   - DT&E: S&TS/AT&L
   - Intel: JS J2/other
4. Revisit T&E data issues/progress in 2004

To Make an Integrated T&E Data Program MØRS
Work...Four Things are Needed

1. Composability Framework
2. Component Services Environment
3. Critical Mass of End-User Content
4. Business Model for Life-Cycle Sustainment

How do we leverage the events to satisfy
M&S VV&A needs?

Composability: taxonomy (having pieces of data to compose)

Component services: tools to find, locate and reuse: assemble and employ

End-user content: Of all stake-holders, the bottom or end-user level of the data: results of end-user flow upward in aggregated and condensed forms to decision makers (end-user buy-in)

Must be EASIER for end-user to adapt T&E data program than it would be to just re-produce the data

Guiding Principles of Data Management

1. Users determine when data is pulled, what data is needed, and the value of that data for their application
   i. User-consumers may post their data need and producer communities are empowered to find ways to meet those needs
   ii. Measures of Effectiveness (MOE) will be related to the consumers' ability to effectively execute their mission
   iii. Communities of Interest (COIs) will determine the necessary "tightness" of data exchange, e.g., from tightly coupled to highly decoupled exchanges

Guiding Principles #1

m◉RS

Implications:

    a. The full range of data being produced (e.g., documents, web pages, databases, system outputs, etc.) will need to be "tagged" with metadata

    b. Sufficient metadata must be available to allow the user to make the determination of whether that data meets their need

    c. COIs will have to incorporate user-consumer feedback in their processes

    d. COI metrics will need to be user focused

    e. Allowing for user ratings and publication of those source ratings may be desirable

    f. User-consumers must have a forum for publicizing their "data wanted" needs

    g. Unmet "data wanted" needs will need to be adjudicated across COIs to assess reasonableness of meeting market demand

    h. Translation and mediation must occur to support less tightly coupled exchanges

    i. COIs will be created and dissolved as needed; users and producers can be members of multiple communities

    j. Individuals will be able to share data in a non-hierarchical way across and within communities

54

Guiding Principles of Data Management

2. Producers must post data such that it is accessible on the network by moving data from internal organizational private space to shared spaces.

   i. Posted data limited only to a community must be justified by "need to protect" rather than assuming that all users must provide "need to know"

   ii. Authorization to access is assumed to be given unless otherwise limited

   iii. COIs must facilitate the ability for data to be discovered and shared

Guiding Principles #2

m🌐RS

- Implications:
  a. Metadata must provide a way to identify security limitations
  b. All metadata for posted data must be registered in a "directory" or catalog so that all user-consumers can be aware of what data is available
  c. Data producing systems must be identified in community catalogs to allow users to quickly find areas of interest
  d. COIs are registered in a "directory" or catalog that is accessible to anyone on the network
  e. COIs that are registered must have a defined scope or charter in order to be accurately characterized in the directory
  f. Strong, funded governance and cross COI facilitation must be in place to ensure that COIs mitigate stovepipes
  g. Incentives must be defined to promote data sharing

Guiding Principles of Data Management **m🌐RS**

3. Producers are the responsible source for defining their data content and providing the metadata.
   i.   Producers of data are the best candidates to describe their data.
   ii. Producers are responsible for updating their data and metadata.
   iii. New producers and new consumers of data are being created all the time.

Guiding Principles #3

- Implications:
    a. A core set of metadata will be defined but communities will be allowed to add metadata specific to their needs
    b. Changes arising from analyses, manipulation or interpretation of data on the network creates new data that can be posted by a new producer
    c. Keeping track of data pedigree (e.g., audit trail) is important to enable the user to determine if the source can be trusted
    d. COIs will need to establish some level of configuration management for change and update of posted data
    e. COIs may want to determine authoritative sources within their community; COIs may want to establish authoritative sources across the enterprise
    f. Adding new postings to a directory must be simple—a self service function
    g. Any new postings must be required to add/update a minimum set of metadata with their posting
    h. Subscribe and alert capabilities must be implemented to allow notification to users of changes to data and other sources of related information

Summary of Guiding Principles

1. Users determine when data is pulled, what data is needed, and the value of that data for their application.
2. Producers must post data such that it is accessible on the network by moving data from internal organizational private space to shared spaces.
3. Producers are the responsible source for defining their data content and providing the metadata.
4. Data Management must be flexible enough to support a wide variety of missions, a changing organizational structure and the diverse needs of the Department's users.
5. Measures must be used to evaluate the success of achieving the DM component of net-centricity, as reflected in these guiding principles.

MORS

Working Group 2 Outbrief

Use of VV&A'd M&S Tools to Support T&E

Test and Evaluation, Modeling and Simulation and
VV&A: Quantifying the Relationship Between Testing
and Simulation

MORS Workshop
Kirtland AFB
15-17 October 2002

Participants

MORS

John Haug (Co-chair)	Jack Keane (Co-chair)
Scott Simpkins (Recorder)	Frank Maestas
Steve Brown	Luis Melicio
Suzanne Beers	Mike Phillips
Sergio De La Reza	Randall Riddle
Kevin Fischer	Roy Rice
Luis Hernandez	Seth Shepherd
Bill Hughes	James Thompson
Yvonne Limas	Brad Warner
Geri Lentz	Philip Whittle

Working Group Objectives

Explore guidance and develop recommendations related to the experience of the T&E community in properly using verified, validated and accredited M&S tools to support both Developmental and Operational T&E processes.

Policies/Guidance

- DoD 5000...
- DoD requirements process
- DMSO recommended practices guide
- Services ACQ/M&S guidance
    - Life cycle support plan
    - ACQ strategy
    - ORD
    - SSP/MSSP
    - TEMP

Experiences

m⊕RS

- Most successful when planned up front in the acquisition process, to include:
  - Intended use of M&S as an integral part of systems engineering effort
  - Use of accepted V&V processes
  - Data collection for V&V and model application
  - Participation by "right" people (consensus building)
  - Use of external V&V services
  - Incentives for use of M&S
- PPBES encourages stove piping - discourages networking

Notes: PPBES creates environment in which saving money equates to losing the money: No mechanism for taking advantage of money saved. Hard for a PM to justify expenditures that support beyond their own program.

Experiences Cont'd

m⊕RS

Finding the right model is difficult

Model maintenance
User training   } Who's responsible?
V&V documentation

Best Practices

MORS

"PLAN UP FRONT"
Start Before ORD is finalized
S.E. taxonomy - trace to requirements
Including funding commitments
Robust MSSP/SSP (including T&E)
- Intended Use of M&S
- V&V activities
- Data collection for V&V and model application
- Participation by "Right" People
- Work/plan with industry (key partner)
- Develop internal procedures
- Use of external V&V services
- Incentives for use of M&S
- Use of business process tools (doors, rational rose)

Best Practices Cont'd

MORS

- Training and application
  - DMSO recommended practices guide
  - Statistical methods/design of experiments
  - Claim/argument/evidence construct
  - MORS website

Recommendations

m🌐RS

- Institute iterative M&S requirements process
- Standup a robust, accessible centrally funded M&S support/assistance (1-800 MODELSRUS)
  - Facilitate the following objectives:
    - Data types
    - Models/information
    - M&S training

Recommendations

m🌐RS

- **Continuity/accountability for PMs**
- **Training enhance M&S curriculum**
  - **DAU**
  - **DTEPI**
  - **MORS**
  - **Other sources**
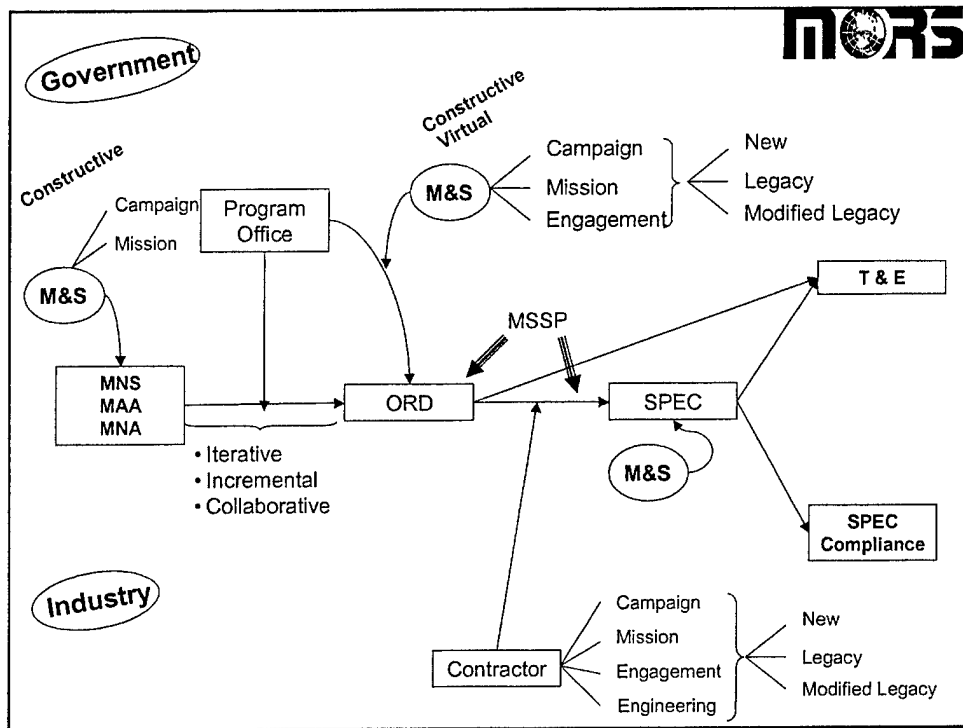
Working Group 3 Outbrief

Accreditation Thought Processes and Issues

Test and Evaluation, Modeling and Simulation and VV&A: Quantifying the Relationship Between Testing and Simulation

MORS Workshop
Kirtland AFB
15-17 October 2002

Working Group 3 Objectives

- Gain insight into current accreditation processes and implementation from the perspectives of the T&E and M&S communities
- Identify best practices for accreditation of M&S in support of T&E based on existing policies and procedures, barriers to implementation and lessons learned
- Identify gaps in policy, procedures and practices and make recommendations on future investments that could benefit the T&E and M&S communities

Outline

- VV&A Planning – Approach
- VV&A Planning – Organization
- Policies
- Data
- Documentation
- Accreditation Approaches
- Resources
- Recommendations

VV&A Planning - Approach

- In successful cases, VV&A was planned upfront, and often integrated as a normal part of the M&S development process
- Plans were variously able or unable to address specific accreditation criteria, required expertise and resource requirements

A plan should be devised to look at models before a test is planned – **model-test-model** (John Haug - ATEC).

There lacks a consistent framework for addressing the various factors that affect VV&A success:

...longer list?

VV&A Planning - Organization

- In general, it was recognized that all stakeholders must be involved in VV&A planning
- Scoping resources and program level risk was difficult
- Although statistical techniques were heavily used, most program risk was qualitative
- Establishing consistent views of the simulation are essential to integrating all the stakeholders

All stakeholders involved means IPT structure. There was often a lack of dedicated effort, proper resources and ability to form a consistent view of the simulation that made VV&A inefficient and/or unsatisfying.

Many projects were able to provide technical, but not program risk, which affected the ability to make a defensible argument for resources. (JSHIP – reallocated program funds from ATEC (who believed in the VV&A) to meet shortfall instead of augmenting.)

It was difficult to formally rectify quantitative and qualitative metrics, inhibiting multidisciplinary communication.

Statistics could not express program risk. (Hollis – the evaluations will be primarily qualitative.) (AIM-9X – exception to qualitative risk.)

Policies

- DoD, Service policies and M&S agency resources were available and used
- There are significant gaps in the implementation of policy
  - Lack of structured examples
  - Lack of accountability to standards
  - Lack of consistent understanding of terminology
- Education and training at the management and action level would be beneficial

General versus specific accreditation: Class accreditation versus application accreditation (ATEC policy)

Each new release of the models need to be accredited for new uses.

The lack of the ability to bridge the gap between policy level concerns and technical achievements and requirements inhibited the implementation of policy.

Structured examples encourage critical thinking about the processes.

Terminology inhibits the ability to construct defensible arguments and capitalize on other efforts (database search in German).

Defensible arguments were not available for resourcing, etc.

DoD and Service policies provided "statutory requirements." DMSO RPG provided technical level information. There were no references sighted on how to bridge the gap between executive level and technical level VV&A. While the consensus of our working group was that there exists no one approach, there exists a core of knowledge that will allow effective and efficient VV&A.

Data

- VV&A data was obtained from existing M&S development processes and from dedicated events
- Most projects have data, but this information is not readily available

Capitalizing on existing VV&A data sources was difficult. Where cross communication between VV&A and the development, configuration management, DT, and/or OT processes was present, the VV&A process was able to capitalize on common data sources. (COATS ...?)

VV&A data comes from different sources which include "real world" data, T&E data and other referents. Justifying dedicated events was difficult. Field tests are/were not designed to support M&S activity (for VV&A). Again, terminology inhibits the ability to recognize VV&A data as program (useful) data.

The relevance of information with respect to VV&A was difficult to determine.

Also, data sources are held proprietary, or not shared.

Diane Burner – Just learning about VV&A, need more upfront experience, too much trial and error.

Need a standard VV&A process/criteria that the entire community can use.

Unified Context – Tie VV&A to a specified project.

73

Documentation

- It is difficult to determine what documentation is available
- It is difficult to determine what to document
- Access is inhibited by:
  - Lack of M&S development documentation
  - Poor documentation implementation
  - Don't have a consistent way to describe the Conceptual Model (CM) (beginning)
- Documentation is under-resourced; resourcing is not reported

The gap between technical and program level information was often bridged by SME opinion, without formal or explicit rationale (another SME couldn't repeat the analysis process).

Accreditation Approaches

- VV&A is not a single process
- There is no single approach to each of the processes
- SMEs are the primary bridge between V&V data and the accreditation decision
- Accrediting for a new intended use, including accrediting distributed simulations and federations, increases reliance on SMEs
- Exemplary accreditation processes:
  - JSHIP – Accreditation Board
  - JWARS – V&V by data taxonomy
  - AIM-9X – Regression testing as an essential element

Different processes apply for different uses of models (Engineering level versus System level):

- Use Process

- V&V process

- CM Process

Management and action level people often viewed VV&A as a burden, rather than an asset.

Resources

- There were no formal methods for estimating VV&A resource requirements
- VV&A is not a integral part of M&S development
- There is no formal method for specifying the intended use
  - Accreditation criteria – must be robustly defined
  - Lack of linkage between intended use and credibility
- Program offices need to account for VV&A cost early in the acquisition process
- The acquisition process lacks program incentives to advance the common good and to encourage critical thinking in M&S

Integral part – Scott Simpkins (JHU/APL) - Good plan not executed.

AIM-9X – MSSP not associated.

Recommendations

m⊕RS

- OSD establish formal reporting requirements for Intended Use as a guide to development, testing, resourcing and VV&A of M&S
- Services establish interim level feedback on VV&A
- OSD and Services must invest in training and education for analysis and management of projects that use M&S
- OSD and Services must define requirements for documentation of M&S development, test data, VV&A and resourcing
- Brief workshop results at the management and action level within OSD and the Services

Intended Use - The intended use of the application defines what the system is going to be accredited for.

**Define intended use: - Need to identify each (measures of merit, how good do each of these need to be):

- What are the outputs?
- Procedures of the operation
- Question to answer
- Inputs

The creation of the intended use should be a "group" project – not just the user of the model.

Final approval of the intended use prior to beginning M&S development is a consumer responsibility!

Interim Feedback necessary, as the process takes shape.

General and/or class accreditation are **not** application accreditations. They provide a mechanism for interim project feedback and identifying commonality among ongoing M&S development processes.

Army class accreditation is a good example.

MORS

Synthesis Group Outbrief

20 November 2002 Presentation to Mr. Hollis, FS

Test and Evaluation, Modeling and Simulation and
VV&A: Quantifying the Relationship Between Testing
and Simulation

**MORS Workshop
Kirtland AFB
15-17 October 2002**

This is the Synthesis Group outbrief from the recent MORS workshop on T&E, modeling and simulation and VV&A.

A minority opinion regarding the findings and conclusions of this brief was submitted by one member of the Synthesis Group, Simone Youngblood, representing DMSO.

Agenda

mⓒRS

- **Progress in VV&A**
- **Successes in VV&A**
- **Challenges and Obstacles to Performing VV&A**
- **Status and Health of VV&A in T&E**
- **Best Practices Using M&S to Support T&E**
- **Recommendations**

These are the topics that I'll cover this morning.

Progress in VV&A

- **Since 1996, DoD has established policies and recommended practices that have provided a basis for understanding VV&A within the T&E community.**
  - DoD 5000...
  - DoD and Services' VV&A policies
  - DoD requirements process
  - DoD Recommended Practices Guide (RPG)
  - Services' acquisition /M&S guidance
    - Life cycle support plan
    - Acquisition strategy
    - Operational Requirements Documents (ORD)
    - Simulation Support Plans (SSP)
    - Test and Evaluation Master Plans

The most significant progress in VV&A has been the publication of policies and planning guides.

These are some of the sources that were cited by the working groups.

Progress in VV&A

mⒶRS

- **The *Army* has had success awarding *class accreditations* that describe the approved general use of a model, which can then be leveraged when the model must be accredited for specific applications within that class.**
- **The *Air Force* (AFSAA) uses a structured process to examine the V&V histories of M&S prior to including them in the *Standard Analysis Toolkit*. A user of the model can then leverage this prior V&V to accredit the model for his or her intended use.**

In addition, some of the Services have developed unique approaches or methods for performing VV&A, most notably, the Army's use of class accreditations.

Successes in VV&A

- **Some programs have overcome technical and programmatic obstacles to plan and perform reasonable VV&A efforts**
  - Hazardous Material Transport and Dispersion Models - Overlapped the figure of merit in red-yellow-green with model results to visualize the validation results
  - Joint Ship Helicopter Integration Program (JSHIP) - The V&V was used to justify a selective accreditation decision
    - The model was used for training/crew certification, but not for flight envelope testing/creation
  - GPS SPIDAR Navigation Accuracy Tool - Used multiple referents for comparison (real world data and the existing model)

Specific successes in VV&A that were briefed at the workshop included these presentations.

Challenges and Obstacles to Performing VV&A

- A perception exists that VV&A is too costly
- A perception exists that VV&A is not integrated in the M&S process and is, therefore, not adequately resourced
- VV&A planning is often conducted without proper scoping up front, resulting in unnecessary work and cost
- VV&A efforts often do not have the right people involved
- Model reuse has not occurred as anticipated

Just as successes were noted, so too were challenges and obstacles to performing VV&A.

Many of these obstacles concerned the costs of doing VV&A.

Status and Health of VV&A in T&E

- **The emphasis since 1996 has been on bureaucratic aspects of VV&A such that the original intent of doing VV&A has been lost**
- **Responsibility for implementing M&S and VV&A rests among people who don't understand the models**
- **There is the perception that some DoD leaders no longer strongly advocate the use of M&S because M&S have not delivered as promised, however, this message has not been shared with lower echelons**
  - Lower echelons continue to use M&S and attempt VV&A because they believe that it is expected
  - Time and energy expended on M&S and VV&A could be better used on testing and other tasks

Marion Williams, FS, asked the Synthesis Group to assess the Status and Health of VV&A in T&E.

We found that an overemphasis has been placed on bureaucratic aspects, such as issuing and reissuing policies and procedures, without providing any substantively new information to those who are trying to perform VV&A.

Verification is more mature as a practice and has a tool set to support it.

Validation is much less mature. Progress is needed to better understand the mathematical and physical underpinnings of a model.

Accreditation methodologies exist. However, applying them with appropriate rigor needs to be matured.

Status and Health of VV&A in T&E

- **Independent V&V has not been realized as originally envisioned**
  - Reliance on V&V conducted by the developer and others who have a vested interest in the model's success
  - The absence of independent V&V has engendered a lack or perceived lack, of objectivity
- **Some programs are avoiding using M&S simply to avoid doing VV&A**
- **The Model-Test-Model paradigm has not been realized as originally anticipated**
- **Few, if any, examples of successful, rigorous validation and accreditation efforts exist for federations or systems of systems**

A rather alarming finding was that some programs were avoiding using M&S simply to avoid having to do VV&A.

An equal concern is that we have been largely unsuccessful in institutionalizing VV&A at the individual model level, and there are expectations of being able to VV&A large federations and systems of systems.

**Best Practices Using M&S
to Support T&E**

1. Understand the problem
2. Determine use of M&S
3. Focus accreditation criteria
4. Scope the V&V
5. Contract to get what you need

Five best practices were identified as a result of the workshop.

It's interesting to note that these were all previously identified as best practices in the 1996 DoD RPG, which has been used with limited success in the intervening years.

Each of the best practices will be further discussed in the following slides.

Best Practices
Understand the Problem

- **Understand the problem/questions so that you can develop sound criteria against which to evaluate possible solutions**
- **This first step is frequently disregarded:**
  - Too hard to do
  - Assumption that the problem is "clearly evident"
  - No real understanding of what the problem is
- **Failure to understand the problem results in:**
  - Unfocused use of M&S
  - Unbounded VV&A
  - Unnecessary expenditure of time and money without meaningful results

Understanding the problem is the first step in the scientific method, but is often not done.

Of course, failure to understand the problem is the root cause of not being able to use M&S effectively.

Best Practices:
Determine Use of M&S

- **The problem may not be VV&A at all, but how we use M&S.**
- **M&S should be used to:**
  - help critical thinking
  - generate hypotheses
  - provide insights
  - perform sensitivity analyses to identify logical consequences of assumptions
  - generate imaginable scenarios
  - visualize results
  - crystallize thinking
  - suggest experiments; the importance of which can only be demonstrated by the use of M&S
  - reduce risk
- **Know up front what you will use M&S for**
  - What part of the problem can be answered by M&S?
  - What requirements do you have that M&S can address?
  - Use M&S appropriately: If the model isn't sufficiently accurate vis-à-vis the test issues, it may lead to flawed information for decisions, such as test planning and execution.

A key point that was made in the workshop was that the problem may not be VV&A at all, but the way in which we use M&S.

Several appropriate uses for M&S were identified and the best practice urges the user to clearly understand what requirements of the problem require the use of M&S for resolution or insights.

A related view that was voiced in the workshop is that the VV&A effort should be focused on reducing program risk. The risk perspective will be discussed further later in this briefing.

Best Practices:
Focus Accreditation Criteria

m⊕RS

- **"The formal process of accreditation should mirror the processes that accompany all research: the process of convincing oneself that the methods and results are reasonable, appropriate, and worth believing to some extent, and of assessing the confidence one has in the results"**
  - Dr. Ernest Seglie and Dr. Patricia Sanders, SIMVAL II, 1992
- **Focus the accreditation on the criteria; focus the VV&A effort on reducing program risk**
- **Accreditation criteria are best developed through collaboration among all stakeholders:**
  - Get System Program Office buy-in on the M&S effort
  - Foster collaboration with testers, analysts and model developers...and key decision makers
  - Build the training simulation first to elicit the user's needs and ideas about "key criteria" before building the system itself. If this is not feasible, think "training" and ultimate use to define key criteria

Even if you know for what you will use M&S, you can't evaluate its suitability or credibility without some criteria against which to assess it.

Establishing sound accreditation criteria that are linked back to the requirements provide the basis for evaluating the model.

Best Practices:
Scope the V&V

- **Focus the accreditation on the criteria as that will properly focus the V&V effort**
- **Plan the V&V effort in advance to use test data to validate the model:**
  - Need to ensure that the data are used correctly
  - Need to document the conditions under which the data were collected

Well-defined accreditation criteria are also needed to scope the V&V effort.

Many V&V efforts become unbounded because accreditation criteria were not defined in advance.

Best Practices: m⊕RS
Contract to Get What You Need

- **Include M&S and VV&A requirements in RFPs to obtain bids and estimated costs**
    - Require that bidders provide a conceptual model that can be used to assess the bidder's understanding of the problem
    - Make documentation of M&S and VV&A deliverables under the contract
    - Competition among bidders will generate alternative models and a broader range of ideas
    - Require that the Government announce what M&S it will use as part of their source selection
    - Ensure that the VV&A team is multidisciplinary and includes experts who understand the mathematical and physical underpinnings of the model
    - Ensure that Subject Matter Experts (SMEs) have relevant and current expertise, that these qualifications are documented, and that the rational of each SME for his or her judgements and recommendations are recorded

Finally, a valuable set of contractual best practices were developed from the workshop.

These establish the use of M&S and the conduct of VV&A as part of the PM's business model.

92

Recommendations to MORS

- **Recommend that MORS explore the possible expansion of the MORS Analyst's Handbook to include case studies that describe how M&S was used to support T&E, and critical reviews that describe the associated VV&A efforts**

The next slides provide specific recommendations out of the workshop.
This recommendation is offered to MORS

- **Recommend that the Service Acquisition Executives work to provide education and establish incentives for program managers that require early planning and execution of M&S, VV&A, and data archiving activities. Incentives may include additional resources, and facilitated progress through the program approval process.**
- **Recommend that the DoD OTAs work with model developers to provide test cases embedded within their software deliveries that can be used to verify that the software is installed correctly, and as an example of how the code can be properly used.**

These recommendations and those on the following slide have been referred to the ad hoc T&E Executive Committee for further consideration.

Other Recommendations (cont.)

- **Recommend that OSD (AT&L) and the Service Acquisition Executives establish a task multidisciplinary group to develop strategies to:**
  - Foster "risk-based management" to the use of M&S and VV&A.
    - Prepare a conceptual framework for determining how V&V should vary with M&S size, type and application
  - Foster a "knowledge management" approach to the use of M&S and VV&A
    - Reinforce need to build a solid knowledge base: it will make future VV&A efforts more efficient and add to better science and engineering.
    - Develop incentives for this
  - Foster a "business case" mindset when building repositories for data as well as VV&A efforts
    - The business model needs to be considered first (who pays for populating it, who maintains it, etc.), rather than worrying about the technical details
  - Develop a conceptual framework to address future concepts, such as transformation, capabilities based requirements, systems of systems as opposed to "eaches" (single systems), etc.

This concludes my brief, Sir.

Backup Slides

MORS

Synthesis Group Objectives and Participants

- **Synthesis Group objectives:**
  - Provide a mechanism to ensure cross-fertilization of ideas among the working groups
  - Integrate and synthesize ideas for the workshop
- **Synthesis Group members:**
  - Dr. Priscilla Glasow, Co-Chair   The MITRE Corporation
  - Ms. Christine Fossett, FS, Co-Chair   GAO
  - Mr. Royce Reiss   AFSAA
  - Dr. Cyrus Staniec   Northrop Grumman IT
  - Dr. Daniel Maxwell   Evidence Based Research, Inc.
  - LtCol Suzanne Beers, USAF, Ph.D.   Space & Missile Systems Center
  - Mr. Patrick McKenna   USSTRATCOM
  - Ms. Simone Youngblood   DMSO/JHUAPL
  - Dr. Frank Mello   DOT&E/LFTE

Three Synthesis Group members attended each of the Tracks A, B and C presentations and participated in each of the Working Groups 1, 2 and 3 discussion sessions. Overall, we found that the Track presentations were very informative and provided good case studies of VV&A. The Working Groups also had good discussions of the issues based on the Track information and their own experiences, and stayed focused on their respective topics.

Successes in VV&A

- **There is evidence of quantitative methods being used for looking at the efficacy of a model for prediction:**
  - Quantifying the Predictive Capability of Computational Models - Provided evidence for, and quantified, the biases and inaccuracies of the model as circumstances changed
  - Equivalence Testing presentation - Reverses the method of hypothesis testing such that the null hypothesis states that two events are not the same, and the alternative hypothesis states that the two events are the same. The method demonstrates the similarity of the model with the real event and provides a quantitative assessment of prediction capability

Other presentations discussed specific quantitative methods for assessing a model's predictive capabilities.

Successes in Using M&S to ∏◉R5 Support T&E

- AIM-9X was an example of a well-conducted VV&A effort that used sound statistical methods (Fisher test, regression testing).
- Tactical UAV IOT&E supplemented input feeds to the Tactical Operations Center with M&S to enrich the threat environment.
- The Air Force Electronic Warfare Evaluation Simulator integrates multiple linked simulations to support tests, and uses sensitivity analysis to support validation.

These are three examples that indicated success in using M&S to support T&E.

Challenges and Obstacles: m⬤RS
The Cost of Doing VV&A

- **A perception exists that VV&A is too costly and difficult to perform given schedule and resource constraints**
- **VV&A is perceived as an expense, not an investment**
- **Examples of programs exist that plan VV&A efforts, but don't execute those plans due to changes in leadership and/or resource constraints**

Cost issues are exacerbated by budgeting processes that do not require Program Managers to report M&S expenditures, much less VV&A costs.

To be viewed as an investment, decision makers must know what VV&A costs and what the benefits are to them.

Challenges and Obstacles: MORS
Lack of Integration

- **Though best practice incorporates VV&A as an integral part of the M&S development process, there is, in practice, still evidence of VV&A as an add-on**
    - The good news: many examples (e.g., the presentations at the workshop) where VV&A is integrated into the M&S process.
    - However, where VV&A is viewed as an add-on, resourcing becomes an issue.
    - Program offices need to account for VV&A cost early in the acquisition process.
    - There were no formal methods for estimating VV&A resource requirements.
    - The acquisition process lacks program incentives to advance the common good and to encourage critical thinking in M&S.

Challenges and Obstacles: MORS
Unscoped VV&A Planning

- **V&V plans are often written before accreditation plans, if the latter are written at all, creating a vicious cycle.**
  - The accreditation plan provides the scope of the V&V effort.
  - The accreditation plan also provides the criteria, or metrics, for assessing whether the model should be accredited or not. These criteria may be qualitative and based on the subjective judgement of the decision maker as to what is sufficient for him or her to make an accreditation decision.
  - Without this scope defined in advance, there is no logical answer to what is enough V&V.
  - Without *a priori* understanding of what is sufficient V&V, costs unnecessarily escalate.
  - Escalated costs add to the commonly-held misperception that VV&A is too costly.

To be successful, VV&A needs to be planned upfront, and integrated as a normal part of the M&S development process. However, several groups noted that the order in which VV&A planning is conducted is often backwards, and results in unnecessary work and cost, as well as a VV&A effort that is difficult to execute.

Challenges and Obstacles: MORS
Not Having the Right People Involved

- **Many M&S are developed such that the mathematical and physical underpinnings of the models are not always evident. Need multidisciplinary teams to develop M&S.**
- **Subject Matter Experts (SMEs) provide one of the easiest, fastest and most inexpensive methods to validate and assess a model preparatory to accreditation, but rarely are the experts' qualifications verified or documented. Need multiple SMEs to provide diversity and balance of views and experience.**

Having the right people involved in a VV&A effort is critical to its success, however, too often models are developed by people with backgrounds in computer science, not mathematics and physics.

Multidisciplinary teams are needed.

Similarly, subject matter experts can be very helpful to a VV&A effort, but only if they have the right expertise.

Need to document the qualifications of SMEs and the rationale of each SME for his or her judgements and recommendations.

But we must recognize that the value of M&S is that it often surfaces counter-intuitive behavior, while SMEs look for behaviors that confirm their intuitions.

Challenges and Obstacles: m⊕RS
Absence of Model Reuse

- **Reuse has not occurred as anticipated for a variety of reasons, including:**
  - No incentive to document code to facilitate understanding by other potential users
  - No incentive to populate repositories due to costs and time required to do so
  - No institutional rewards for following directives
  - Reluctance to share information, models and data
  - Data obtained from tests and training exercises is generated, but not used to validate M&S
  - Legal restrictions of proprietary models
  - M&S are stovepiped within the Services and programs, thereby creating obstacles to interoperability and reuse
  - Hesitation to take on VV&A for M&S that have no prior V&V or are not documented.

- A major challenge to VV&A developing as anticipated by the early SIMVAL workshops has been the absence of reuse of M&S.
- This has occurred for several reasons
  - Reasons listed above
- Further,
  - Most projects have data, but this information is not readily available
  - It is difficult to determine what documentation is available
  - It is difficult to determine what to document.
  - Access is inhibited by:
    - Lack of M&S development documentation
    - Poor documentation implementation
    - Don't have a consistent way to describe the Conceptual Model
    - Documentation is under-resourced; resourcing is not reported

104

**Example –
DOT&E Pilot Projects on M&S**

MORS

- **Purpose of pilot program was to develop a methodology that could be used to determine how to better use M&S in support of acquisition programs**
- **Approach:**
  - Characterize data needed to support acquisition decision making
  - Categorize potential opportunities for M&S to support those acquisition decisions
  - Describe the risks and benefits of using M&S-generated data to support these acquisition decisions
  - Determine whether findings can be generalized into methodological approach

An example in which this way of thinking was implemented in practice was given by Dr. Anne Hillegas.

Other Observations

- **Maxwell white paper on the impact of transformation and experimentation on VV&A**

Although outside the scope of this workshop, the topic engendered thoughts on other aspects of VV&A.

For example, Dr. Dan Maxwell of Evidence Based Research submitted a white paper that considers the impact of transformation and experimentation on VV&A.

Memorable Quotes **MORS**

- **"Credibility is an independent variable."**
  - *Dr. Ernie Seglie*
- **"An unfunded requirement is just an hallucination."**
  - *Dr. Roy Rice, PP*
- **" A fool with a tool is still a fool."**
  - *Lt Col Seth Shepherd, USAF*
- **"If you build a tool that even idiots can use, then rest assured, idiots will use it."**
  - *Wilbur Payne, FS*

No workshop would be complete without a few memorable quotes. Here are some that were generated (or repeated) at our workshop.

Synthesis of Working Group Findings

The working groups were provided with a set of templates to use as a guide for their outbriefs.

The following slides examine the information collected through that process and provide additional insights into the subject.

VV&A Planning - Approach

- **WG 2 - M&S to Support T&E**
  - Finding the right model is difficult
    - Model maintenance
    - User training
    - V&V documentation
  - Who is responsible
    - For model maintenance
    - For providing user training
    - For documenting the model

WG 2 was tasked with examining the use of M&S to support T&E. When planning a VV&A effort, they noted that....

VV&A Planning - Approach

- **WG 3 - Accreditation Processes**
  - In successful cases, VV&A was planned upfront, and often integrated as a normal part of the M&S development process
  - Plans were variously able or unable to address specific accreditation criteria, required expertise and resource requirements
  - VV&A is not a single process
  - There is no single approach to each of the processes
  - There is no formal method for specifying the Intended Use
    - Accreditation criteria – must be robustly defined
    - Lack of linkage between Intended Use and credibility

In contrast, WG 3 looked at accreditation processes.

They noted somewhat different considerations when planning a VV&A effort...

VV&A Planning
- Organizational Considerations

m⊕RS

- **WG 3 - Accreditation Processes**
  - In general, all stakeholders must be involved in VV&A planning.
  - Scoping resources and program level risk is difficult.
  - Establishing consistent views of the simulation is essential to integrating all stakeholders.

WG 3 also commented on some of the organizational issues that surround the planning and execution of a VV&A efforts.

VV&A Policies - Help or Hinder?

- **WG 3 - Accreditation Processes**
  - There are significant gaps in the implementation of policy
    - Lack of structured examples
    - Lack of accountability to standards
    - Lack of consistent understanding of terminology
  - Education and training at the management and action level would be beneficial.

WG 3 also provided some thoughts about the limitations of how policies are implemented and what might be done to improve this situation.

VV&A Data Sources ꞁꞁ🌐ꓳꓢ

- **WG 3 - Accreditation Processes**
    - VV&A data is obtained from existing M&S development processes and from dedicated events.
    - Most projects have data, but this information is not readily available to others.

The working groups were also asked to discuss how data are used to VV&A models.

WG 3 noted that...

Primary Approach

mโฉาร

- **WG 2 - M&S to Support T&E**
  - Most successful when planned up front in the acquisition process, and includes:
    - Intended use of M&S as an integral part of the systems engineering effort
    - Use of accepted V&V processes
    - Data collection for V&V and model application
    - Participation by the "right" people (consensus building)
    - Use of external V&V services (outsourcing)
    - Incentives for using M&S
  - PPBES encourages stovepiping - discourages networking

When discussing the primary approach used to plan and conduct a VV&A effort, WG 2 noted that...

Primary Approach

- **WG 3 - Accreditation Processes**
  - SMEs are the primary bridge between V&V data and the accreditation decision.
  - Accrediting for a new intended use, including accrediting distributed simulations and federations, increases reliance on SMEs.

WG 3 also provided their observations about primary approaches for accreditation...

Documenting VV&A

- **WG 3 - Accreditation Processes**
  - It is difficult to determine what documentation is available.
  - It is difficult to determine what to document.
  - Access is inhibited by:
    - Lack of M&S development documentation
    - Poor documentation implementation
    - No consistent way to describe the Conceptual Model
  - Documentation is under-resourced; resourcing is not reported.

A key topic was that of documenting VV&A efforts.

WG 3 noted that...

Recommended Best Practices

- **WG 2 - M&S to Support T&E**
  - Plan up front
    - Start before the ORD is finalized
    - Systems engineering taxonomy - trace to requirements
    - Include funding commitments
    - Robust SSP/MSSP (include T&E)
  - Know the intended use of the model
  - Plan the necessary V&V activities
  - Plan the data collection for V&V and the model application
  - Ensure participation by the "right" people
  - Plan and work the effort with industry as a key partner

WG 2 provided many excellent recommended best practices that new practitioners of VV&A might find useful.

First,...

Recommended Best Practices **MⒺRS**

- **WG 2 - M&S to Support T&E**
  - Develop sound internal procedures
  - Use external V&V services
  - Establish incentives for using M&S
  - Use business process tools, such as DOORS and Rational Rose
  - Use existing resources to train and apply M&S in support of T&E, including
    - DoD Recommended Practices Guide
    - Statistical methods/Design of experiments
    - Claim/argument/evidence construct
    - MORS website

- **WG 1 - Data**
    - Make T&E data more available for M&S applications
    - Support a structured data program
        - Taxonomy
        - Accessibility
        - Business Case
        - Culture
    - Identify potential control centers for T&E data
        - OT&E: DOT&E
        - DT&E: S&TS/AT&L
        - Intel: JS J2/other
    - Revisit T&E data issues/progress in 2004

The WG strongly recommends implementing a structured T&E data program with a senior advocate in the OSD Staff. The Structured program should be set up on a limited basis to develop the logical taxonomy and address the resource and cultural impediments to reuse and sharing of data. This approach would also require a senior sponsor to advocate the reuse of T&E data for VV&A and other M&S uses.

WG 1 recommended a follow-on MORS meeting after two years to determine what progress had been made in the reuse of T&E data.

WG 1 observed that there is a need for a joint executive agent

Before you can go DoD, you must have a lower level traction point

If there is no hammer: it won't get done

Need a carrot and a stick

Any proposal for a new program (regardless of value or need) is going to be a hard sell, especially given the move to get AT&L separated from being an acquisition agent

Don't shoot ourselves in the foot

No money: PM-SPO…

AT&L taxes all programs… and now we are going to tax them more to do data sharing

Get 4 stovepipes together:

1. Find out what/how database repository will work

2. Intel/DARPA data

3. Certified data: current/good data

4. M&S in T&E are only tools

Recommendations

• WG 2 - M&S to Support T&E

- Institute an iterative M&S requirements process
- Stand up a robust, accessible, centrally-funded M&S support/assistance (I-800-MODELS-R-US) to facilitate the following objectives:
  - Data types
  - Models/information
  - M&S training
- Continuity/accountability for PMs
- Training to enhance M&S curricula at
  - Defense Acquisition University
  - Defense T&E Professionals Institute (DTEPI)
  - MORS

Recommendations

m⊕RS

- ## WG 3 - Accreditation Processes
  - OSD establish formal reporting requirements for Intended Use as a guide to development, testing, resourcing and VV&A of M&S
  - Services establish interim level feedback on VV&A
  - OSD and Services must invest in training and education for analysis and management of projects that use M&S
  - OSD and Services must define requirements for documentation of M&S development, test data, VV&A and resourcing
  - Brief workshop results at the management and action level within OSD and the Services

Test and Evaluation, Modeling and Simulation and VV&A: Quantifying the
Relationship Between Testing and Simulation
15-17 October 2002
The Energy Training Complex, Kirtland AFB
Albuquerque, New Mexico

Acronyms

AAV	Amphibious Assault Vehicle
ACQ	Acquisition
ADMAS	Air Defense Models and Simulation
AFOTEC	Air Force Operational Test & Evaluation Center
AFSAA	Air Force Studies and Analyses Agency
AIM-9X	Sidewinder Missile
AMRAAM	Advanced, Medium-Range, Air-to-Air Missile
AMSAA	US Army Materiel Systems Analysis Activity
AoA	Analysis of Alternatives
AT&L	Acquisition Technology and Logistics (OSD)
ATEC	Army Test and Evaluation Command
CDEC	Combat Development Experimental Command
CFD	Computational Fluid Dynamics
CM	Conceptual Model
CNR	Chief of Naval Research
COI	Community Of Interest
COTS	Commercial Off The Shelf
CTP	Critical Technical Parameters
DAB	Defense Acquisition Board
DARPA	Defense Advanced Research Projects Agency
DAU	Defense Acquisition University
DIA	Defense Intelligence Agency
DM	Decision Maker
DMSO	Defense Modeling and Simulation Office
DoD	Department of Defense
DOT&E	Director, Operational Test and Evaluation
DT	Development Testing
DT&E	Developmental Test and Evaluation
DTEPI	Defense Test and Evaluation Professionals Institute
DTIC	Defense Technical Information Center
DUSA(OR)	Deputy Under Secretary of the Army (Operations Research)
ESAMS	Enhanced Surface-to-Air Missile Simulation
FCS	Future Combat Systems

GAO	Government Accounting Office
IACs	Information Analysis Centers
IOT&E	Initial Operational Test and Evaluation
IPT	Integrated Product Team
IR	Infrared
IV&V	Independent Verification and Validation
J2	Joint Staff (Intelligence Directorate)
JHU/APL	Johns Hopkins University/Applied Physics Lab
JMASS	Joint Modeling and Simulation System
JRTC	Joint Readiness Training Center
JS	Joint Staff
JSF	Joint Strike Fighter
JSHIP	Joint Ship Helicopter Integration Program
JSIMS	Joint Simulation System
JWARS	Joint Warfare Analysis and Requirements System
LRIP	Low Rate Initial Production
M&S	Modeling and Simulation
MAA	Mission Area Assessment
MIT	Massachusetts Institute of Technology
MNA	Mission Needs Analysis
MNS	Mission Needs Statement
MOA	Memorandum of Agreement
MOE	Measures of Effectiveness
MORS	Military Operations Research Society
MSIAC	Modeling and Simulation Information Analysis Center
MSRR	Modeling and Simulation Resource Repository
MSSP	Modeling and Simulation Support Plan
NSC	Natick Soldier Center
OA	Operational Assessment
OPTEVFOR	Operational Test and Evaluation Force
ORD	Operational Requirements Document
OSD	Office of the Secretary of Defense
OT	Operational Test
OT&E	Operational Test and Evaluation
OTA	Operational Test Agency
PEO	Program Executive Office
PM	Program Manager
PP	MORS, Past President
PPBES	Planning, Programming, Budgeting and Execution System
RF	Radio Frequency
RFP	Request For Proposal
RPG	Recommended Practices Guide
S&TS	Strategic and Tactical Systems (OSD)
SIMVAL	Simulation Validation
SME	Subject Matter Expert
SPEC	System Specifications

SSP	Simulation Support Plan
T&E	Test and Evaluation
TEMP	Test and Evaluation Master Plans
TMAP	Threat Modeling and Analysis Program
TOW	Tube launched, Optically tracked, Wire guided
UAV	Unmanned Aerial Vehicles
USA	United States Army
USAF	United States Air Force
USA-PM/OTA	United States Army-Program Manager/ Operational Test Authority
USMC	United States Marine Corps
USN	United States Navy
V&V	Verification and Validation
VV&A	Verification, Validation and Accreditation

Test and Evaluation, Modeling and Simulation and VV&A: Quantifying the Relationship Between Testing and Simulation
15-17 October 2002
The Energy Training Complex, Kirtland AFB
Albuquerque, New Mexico

Terms of Reference

Background

As the use of models continues to become more prevalent in system design, expectations are that they also will become a more central tool supporting system development and test and evaluation. In many cases, models or simulations could be expected to become a primary evaluation tool for addressing Operational Requirement Documents (ORDs) and Critical Technical Parameters (CTPs), and for resolving Critical Operational Issues (COIs).

With this increased reliance on M&S in systems acquisition, VV&A of M&S becomes even more critical. While there have been considerable advances in M&S technology, there has been limited comprehensive research and review of VV&A methodology. Most VV&A is accomplished with Subject Matter Experts (SMEs) who review simulation predictions and test results, and provide a subjective opinion on the adequacy of the model. There are few examples of an objective process or quantitative criterion that allows the decision maker to fully understand why the model is or isn't good enough for the particular application. Furthermore, in many instances, there is inadequate definition of the intended application of the model – a critical element in any accreditation.

While SMEs play an important role in the accreditation process, test and evaluation data resulting from the use of real equipment and systems should provide a major source of data for V&V of M&S. Statistical tests comparing test results and model results, and perhaps even some overall measure of "goodness of fit", should be able to provide the analyst with a means of quantifying the validity of model results and assessing the degree to which models can support weapon system evaluation.

An issue of concern to the analyst and decision-maker in using models to supplement field-testing is the lack of credibility of the model in predicting outcomes outside its validated domain space. While testing all aspects of a model is not typically economically feasible, validating a model in the center of the

envelope and using it to predict performance at the edge of the envelope is not practical or valid.

1. Objectives and Goals

This workshop has three interrelated, though distinct, objectives. These are to explore guidance and develop recommendations related to the:
1. Experience of the M&S community in the use of test and evaluation (T&E) data to support the VV&A of M&S.
2. Experience of the T&E community in properly using verified, validated, and accredited M&S to support both Developmental and Operational T&E processes.
3. Experience of the T&E and M&S communities in Accreditation thought processes and implementation.

The overall goals of this workshop are to:
- Develop a set of recommended Best Practices that will be focused on VV&A methodologies for use in T&E derived from collecting and synthesizing:
  1. Existing policies and procedures,
  2. Previous or known barriers to implementation
  3. Lessons learned in these two areas.
- Determine gaps in the associated policies, procedures and practices that would be fruitful areas of future investment that would benefit the T&E and/or M&S communities.

The meeting will highlight potential ways of:
- Applying specific VV&A methodologies to T&E problems, and
- Incorporating the results into DoD/Service & Agency T&E VV&A procedures.

2. Approach

The proposed MORS Workshop, "**Test & Evaluation, Modeling & Simulation, and VV&A: Quantifying the Relationship between Testing and Simulation,**" will be organized by three general topic areas: T&E for M&S V&V, M&S Support to T&E, and Accreditation Processes.

- Senior DoD T&E government representatives will present briefings on the first day of the workshop. The speakers will highlight issues in T&E and VV&A from their respective and unique perspectives.

Workshop Co-Chairs: Dr. Marion Williams, Ms. Annie Patenaude

Opening remarks:
M&S expectations and the challenge of VV&A: Mr. Walt Hollis, accepted

Perspectives:
Army OT&E: Mr. Brian Barr, accepted
Navy OT&E: Mr. Steve Whitehead, accepted
Air Force OT&E: Dr. Marion Williams, accepted
DOT&E: Dr. Ernest Seglie, accepted

- Attendees will participate in working groups to discuss issues related to the workshop goals and objectives and hear case studies from which they will develop a set of proposed best practices.

Working Groups: To support the workshop goals and objectives, case studies will be sought in model applications at all levels. The working groups will attend briefings of the case studies and focus in three areas:

Working Group 1   Use of T&E data to support M&S VV&A
- Policy & Procedures
  - Includes Basis of Confidence issues
  - Includes discussion of implementation difficulties
- Use of T&E data to support M&S VV&A: Experiences and Lessons Learned

Working Group 2   Use of VV&A'd M&S tools to support T&E
- Policy & Procedures
  - Includes models for training, development and predictive applications
  - Includes discussion of barriers to implementation
- Use of VV&A'd M&S tools to support T&E: Experiences and Lessons Learned
  - Includes both DT and OT perspectives

Working Group 3   Accreditation Thought Processes and Issues
- Includes discussion of key information needed by decision makers

The case studies will be solicited from those using various types of models and analysis to support the Test and Evaluation.

Campaign Level (JWARS)

Mission Level (SUPPRESSOR)

Engagement Level (ESAMS/JMASS)

Engineering level (CFD)

Physics level (Hydrocodes)

How do you make sure those factors are represented to an adequate level of fidelity for that application?

How do you determine the important factors for an application?
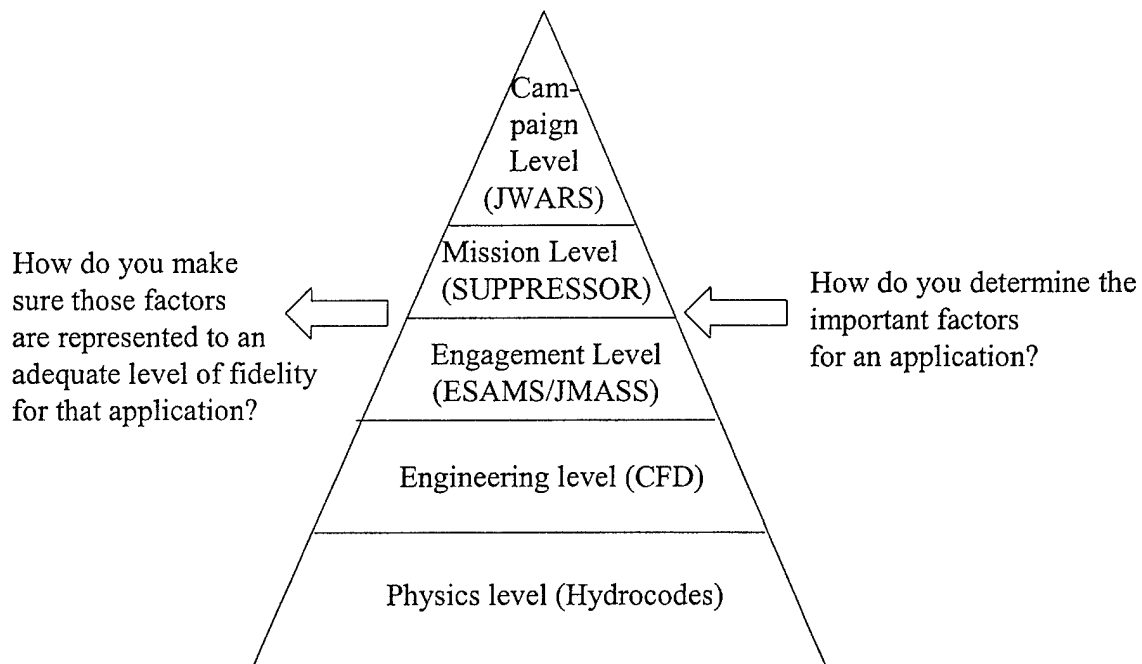
Synthesis Group Co-Chairs: Priscilla Glasow, Chris Fossett

Publications Bulldog: Lana McGlynn

3. **Working Group Chairs and Co-Chairs.** The chairs and co-chairs of the working groups, along with their contact information, are identified in Table 1 below.

Table 1 Working Group Chairs and Co-Chairs

Working Group 1	Maj Todd McDonald	AFOTEC	505-853-7757
Working Group 1	Fred Hartman	JSIMS	703-604-7234
Working Group 2	John Haug	ATEC	410-278-1275
Working Group 2	Jack Keane	JHU/APL	240-228-8886
Working Group 3	LT Alex Hoover	OPTEVFOR	757-444-5564
Working Group 3	Dennis Lester	JGPSCE	505 853-7395
Track A Chair	Greg Keethler	HQ AFSPC	719-554-7744
Track B Chair	Kevin Fischer	DCS Corp	703-697-7703
Track C Chair	Charles Walters	OSD JT&E	703-681-4024 x 171

4. **Meeting Location.**
The workshop will be held in Albuquerque, NM, 15-17 October 2002 at Kirtland AFB. The MORS Hotel will be the Radisson Hotel & Conference Center, 505-888-3311.

5. Agenda

Table 2 highlights the top-level schedule for the three-day workshop.

Table 2 Top-Level Workshop Schedule

	Tuesday 15 Oct 2002	Wednesday 16 Oct 2002	Thursday 17 Oct 2002
Morning	Plenary Service & Agency Speakers	Case Study Presentations	Case Study Presentations
Lunch	Lunch	Lunch	Lunch
Afternoon	Service & Agency Speakers/ Leadership Panel	Case Study Presentations	WG Outbriefs Synthesis Chair Discussion & Actions
Break			
	WG Break-Out	WG Break Out	Recommendations to Panel
Evening	Mixer	WG Chair Hot Wash	WG Chair Documentation

6. Attendees

The working group chairs will be key OTA representatives from each of the Services and Agencies. Co-chairs will be MORS members from T&E and analysis communities. Invitees will be from DoD, Industry, and Academia with interest in T&E M&S use.

7. Products

In addition to the workshop products of *PHALANX* article and briefs to MORS Sponsors, the goals of this workshop support bringing a credible synthesis of recommendations to the practicing community. The products from the working groups will be synthesized into out-briefs to the T&E Leads and Commanders in the Services and Agencies, with guidance and recommendations for T&E VV&A methodologies. These products will provide a draft set of Best Practices and identify existing gaps in VV&A and T&E policies, procedures and practices that would be fruitful areas for future investment.

Additionally, the Working Group Co-Chairs will present a synopsis of the workshop, its recommendations, and resulting actions during a special session of the MORS Symposium in June 2003.

Organizing Committee

Name	Phone	e-mail address	Organization
Brian Engler	703 933-9070	brian@mors.org	MORS
Natalie Kelly	703 933-9070	natalie@mors.org	MORS
Ted Smyth	240 228-6342	Ted.smyth@jhuapl.edu	MORS
Dr. Tom Allen	703 578-2773	tallen@ida.org	MORS
Dr. Roy Rice	256 726-2038	Roy.rice@tbe.com	MORS
Annie Patenaude	703 556-1378	apatenaude@northropgrumman.com	MORS
Jack Marriott	(703)633-8300 x4672	jmarriott@northropgrumman.com	MORS
Dick Wiles	843 342-2076	dickwiles@hargray.com	MORS
Dr Marion Williams	505 846-0607	marion.williams@afotec.af.mil	AFOTEC
Lana McGlynn	703 697-0367	Lana.mcglynn@hqda.army.mil	ODUSA(OR)
Chris Fossett	202 512-2956	fossettc@gao.gov	GAO
Priscilla Glasow	703 883-6931	pglasow@mitre.org	MITRE
Pat McKenna	402 294-1654	mckennap@stratcom.af.mil	STRATCOM
Suzanne Beers	719 556-2829	Suzanne.beers@cisf.af.mil	SMC/CWSN
Dean Free	703 801-1798	dfree@anteon.com	Anteon
John Haug	410 278-1275	haugjohn@atec.army.mil	ATEC
LtCol Gary Barmore	703 697-4087	Gary.barmore@bmdo.osd.mil	MDA/TEM
Fred Hartman	703 604-7234	Frederick.Hartman@saalt.army.mil	DMSO
Simone Youngblood	703 824-3436	syoungblood@dmso.mil	DMSO
Brian Hall	757 444-5546	hallb@cotf.navy.mil	OPTEVFOR
COL Generazio	703 697-0366	Hoa.generazio@hqda.army.mil	DUSA(OR)
Jack Keane	240 228-8886	Jack.keane@jhuapl.edu	JHU/APL
Dennis Lester	505 853-7395	Dennis.lester@afotec.af.mil	JGPSCE
Tom Tanner	540 663-2137	ttanner@dv.synthetics.com	ACS-Synthetics
Alex Hoover	757 444-5546	hoovera@cotf.navy.mil	COTF
Greg Keethler	719 554-7744	Greg.keethler@peterson.af.mil	HQAFSPC
Kevin Fischer	703-697-7703	Kevin.fischer-contractor@bmdo.osd.mil	DCS Corp
Charles Walters	703-681-4024 x 171	Charles.walters@osd.mil	OSD JT&E

October 4, 2002